



TECHNOLOGY AND APPLICATIONS SERIES

SCOTT GLEASON • DEMOZ GEBRE-EGZIABHER
editors



GNSS

APPLICATIONS AND METHODS



DVD
INCLUDED

GNSS Applications and Methods

DISCLAIMER OF WARRANTY

The technical descriptions, procedures, and computer programs in this book have been developed with the greatest of care and they have been useful to the author in a broad range of applications; however, they are provided as is, without warranty of any kind. Artech House, Inc. and the author and editors of the book titled *GNSS Applications and Methods* make no warranties, expressed or implied, that the equations, programs, and procedures in this book or its associated software are free of error, or are consistent with any particular standard of merchantability, or will meet your requirements for any particular application. They should not be relied upon for solving a problem whose incorrect solution could result in injury to a person or loss of property. Any use of the programs or procedures in such a manner is at the user's own risk. The editors, author, and publisher disclaim all liability for direct, incidental, or consequent damages resulting from use of the programs or procedures in this book or the associated software.

For a listing of recent titles in the *Artech House Technology and Applications Series*, turn to the back of this book.

GNSS Applications and Methods

Scott Gleason
Demos Gebre-Egziabher

Editors



**ARTECH
HOUSE**

BOSTON | LONDON
artechhouse.com

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the U.S. Library of Congress.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library.

ISBN-13: 978-1-59693-329-3

Cover design by Igor Valdman

© 2009 Scott Gleason and Demoz Gebre-Egziabher

All rights reserved.

Artech House

685 Canton Street

Norwood, MA 02062

Printed and bound in the United States of America. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Artech House cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

10 9 8 7 6 5 4 3 2 1

Disclaimer:

This eBook does not include the ancillary media that was packaged with the original printed version of the book.

Contents

Preface	xv
CHAPTER 1	
Global Navigation Satellite Systems: Present and Future	1
1.1 Introduction	1
1.1.1 Current and Planned GNSS Constellations	2
1.1.2 GNSS User Architectures	3
1.1.3 Current GNSS Applications	6
1.1.4 Positioning Performance Measures	8
1.2 GNSS Signal Improvements	9
1.2.1 Additional GPS Frequencies	10
1.2.2 Higher Accuracy Ranging	10
1.2.3 Longer Ranging Codes	11
1.2.4 Higher Transmit Power Levels	11
1.3 Advanced Receiver Technology	12
1.3.1 Conventional Receivers	12
1.3.2 FPGA-Based Receivers	13
1.3.3 Software-Defined GNSS Receivers	14
1.4 Road Map: How To Use This Book	14
1.5 Further Reading	20
References	21
CHAPTER 2	
GNSS Signal Acquisition and Tracking	23
2.1 Introduction	23
2.2 GNSS Signal Background	23
2.2.1 BOC Signal Modulation	25
2.2.2 PRN Codes	26
2.3 Searching for PSK Signals	28
2.4 Tracking PSK Signals	34
2.4.1 Phase-Locked Loop (PLL)	34
2.4.2 Frequency-Locked Loop (FLL)	36
2.4.3 Delay-Locked Loop (DLL)	37
2.5 Searching for BOC Signals	39

2.6	Tracking BOC Signals	42
2.6.1	BOC Tracking Using a Single Sideband (SSB)	44
2.6.2	BOC Tracking with Multiple-Gate Discriminators (MGD)	44
2.6.3	BOC Tracking with the Bump-Jumping (BJ) Algorithm	46
2.6.4	BOC Tracking with the Dual Estimator (DE)	48
	References	53

CHAPTER 3

	GNSS Navigation: Estimating Position, Velocity, and Time	55
3.1	Overview	55
3.2	Position, Velocity, and Time (PVT) Estimation	56
3.2.1	Estimating Receiver Position and Clock Bias	56
3.2.2	Impact of Ionosphere Errors	62
3.2.3	Impact of Satellite-User Geometry (DOP)	63
3.2.4	Estimating Receiver Velocity and Clock Drift	64
3.2.5	Estimating Time	66
3.2.6	PVT Estimation Using an Extended Kalman Filter (EKF)	67
3.2.7	Enhanced Accuracy via Carrier Phase Positioning	67
3.2.8	Error Sources	67
3.3	GNSS Simulator	69
3.3.1	GNSS Simulator Measurement Details	69
3.3.2	GNSS Simulator Interface Files	71
3.3.3	Postprocessing GNSS Simulator Output Files	73
3.4	GNSS Simulator Examples	74
3.4.1	Example 1: Simple Navigation	74
3.4.2	Example 2: Traveling Between Destinations	75
3.4.3	Example 3: Waypoint Navigation Using FlightGear	77
3.4.4	Example 4: Dual-Frequency Calculation	80
3.4.5	Example 5: Adding Galileo Satellites	82
3.4.6	Example 6: Spacecraft-Based Receiver	84
3.5	Summary	84
3.6	Programs and Tools Provided on the DVD	86
	References	86

CHAPTER 4

	Differential GNSS: Accuracy and Integrity	87
4.1	Introduction to DGNSS	87
4.2	Fundamentals of Differential GNSS	87
4.2.1	Error Sources and Degree of Spatial Correlation	89
4.2.2	Local Versus Regional DGNSS Corrections and DGNSS Networks	93
4.2.3	Means of Distributing DGNSS Corrections	94
4.2.4	Managing the Latency of DGNSS Corrections	96
4.3	DGNSS Integrity Threats and Mitigations	97
4.3.1	Integrity Threats from GNSS Faults	98

4.3.2 Integrity Threats from DGNSS System Faults	108
4.3.3 Integrity Threats from Signal Propagation Anomalies	109
4.4 Summary	114
4.5 Data Provided on the DVD	115
References	115

CHAPTER 5

A GPS Software Receiver	121
5.1 Introduction and Background	121
5.2 License, Development Environments, and Tools	122
5.2.1 License	122
5.2.2 GNU/Linux	122
5.2.3 Microsoft Windows	123
5.2.4 Apple Mac OS X	123
5.2.5 Displaying the Receiver Output	123
5.3 Example Data Sets	123
5.3.1 Data Set 1	124
5.3.2 Data Set 2, for Use with WAAS Corrections Data	124
5.4 Using the fastgps Software Receiver	124
5.4.1 Configuration File	124
5.4.2 Output Files	130
5.5 fastgps Software Receiver Architecture	131
5.5.1 Timing and Clock Management	132
5.5.2 Main Processing Loop	133
5.5.3 Acquisition	133
5.5.4 Tracking	136
5.5.5 Navigation	142
5.6 Suggested Future Improvements	145
5.7 Further Reading	146
References	146

CHAPTER 6

Integration of GNSS and INS: Part 1	149
6.1 Introduction	149
6.2 Inertial Navigation	150
6.2.1 Inertial Sensors	150
6.2.2 Coordinate Frames	151
6.2.3 Mechanization Equations	152
6.2.4 System Initialization	157
6.2.5 INS Error Model	157
6.3 GNSS/INS Integration Concepts	159
6.3.1 Motivation for GNSS/INS Integration	159
6.3.2 Integration Architecture Overview	160
6.3.3 Loose GNSS/INS Integration	160
6.3.4 Tight GNSS/INS Integration	162
6.3.5 Deep GNSS/INS Integration	164

6.4	Filtering/Estimation Algorithms	165
6.4.1	Overview of Extended Kalman Filter (EKF) for GNSS/INS	165
6.4.2	Time Evolution of a GNSS/INS System	168
6.5	GNSS/INS Integration Implementation	169
6.5.1	IMU Sensor Error Models	169
6.5.2	GNSS/INS Integration: Step-by-Step	172
6.6	Practical Considerations	172
6.6.1	Lever Arm	173
6.6.2	Timing Requirements	173
6.7	Summary and Further Reading	174
	References	175

CHAPTER 7

	Integration of GNS and INS: Part 2	177
7.1	Introduction	177
7.2	Case Study 1: Low-Cost GNSS/INS Integrated Navigator	177
7.3	Case Study 2: Vehicle Sideslip Estimation	181
7.3.1	Motivation	181
7.3.2	Observability	184
7.4	Case Study 3: INS To Aid High-Accuracy GNSS	186
7.4.1	GNSS Ambiguity-Resolution Overview	187
7.4.2	Benefits of INS to Ambiguity Resolution	188
7.5	Software Examples	189
	References	189

CHAPTER 8

	Integrated LADAR, INS, and GNSS Navigation	191
8.1	Introduction	191
8.2	LADAR-Based TERRAIN Integration Methodology	192
8.3	LADAR-Based Terrain-Referenced Position Estimation	196
8.3.1	Position Estimate and SSE Surface	196
8.3.2	Exhaustive Grid Search	198
8.3.3	Gradient-Based Search	200
8.4	Estimation of Inertial Velocity Error	202
8.5	Case Studies of TERRAIN System Performance	202
8.5.1	Case Study I—General Positioning System	202
8.5.2	Case Study II—Precision Approach Guidance System	205
	References	210

CHAPTER 9

	Combining GNSS with RF Systems	211
9.1	Location System Alternatives	211
9.2	RF Location Types and Classifications	213
9.2.1	Location by Proximity	214
9.2.2	Location by Radio Direction Finding (DF) and Angle of Arrival (AOA)	217

9.2.3	Location Using Doppler Frequency	219
9.2.4	Location Estimation Using Signal Strength	221
9.2.5	Location Using Time, Phase, and Differential Timing of Arrival (TOA, POA, and TDOA)	223
9.3	Estimation Methods	226
9.3.1	Deterministic Estimation Using Triangulation	226
9.3.2	Deterministic Estimation Using Nearest Neighbor	229
9.3.3	Nonranging-Based Location Estimation	231
9.3.4	Probabilistic Estimation Using Centroid/Center of Mass	232
9.3.5	Bayesian State Estimation	232
9.4	Integration Methods	234
9.4.1	Least-Squares Integration	234
9.4.2	Kalman Filter Integration	235
9.4.3	Contextual Processing	235
9.5	Example Systems	235
9.5.1	Pseudolites	235
9.5.2	Synchrolites	237
9.5.3	Self-Synchronizing Networks	237
9.5.4	GPS and Relative Navigation	238
9.5.5	TV-Based Location	238
9.5.6	Integration of Cellular Location Systems and GNSS	239
9.6	Examples Included on the DVD	240
9.6.1	RF Antennas	240
9.6.2	Doppler Calculations	240
9.6.3	K-Nearest Neighbor Plot	241
9.7	Further Reading	241
	References	241

CHAPTER 10

Aviation Applications	245
10.1 Introduction	245
10.2 Classes of Aviation Augmentation Systems	245
10.3 Benefits of GPS and Augmentations to Aviation Users	247
10.3.1 Oceanic Flight	247
10.3.2 Overland Flight: En Route, Terminal, and Nonprecision Approach	248
10.3.3 Precision Approach and Landing	248
10.4 Future of GNSS Navigation in Aviation	249
10.4.1 GNSS Modernization	249
10.4.2 Next-Generation Air Traffic Management System (NextGen)	251
10.4.3 Backup Navigation Capabilities for Aviation	251
10.5 Functionality of Aviation Augmentation Systems	252
10.5.1 Augmentation System Performance Requirements	252
10.5.2 Error Bounding Under Nominal Conditions	253
10.5.3 Error Bounding Under Anomalous Conditions	257
10.5.4 Monitoring	261

10.6 Conclusion	264
10.7 Further Reading	265
References	265

CHAPTER 11

Integrated GNSS and Loran Systems	269
11.1 Introduction	269
11.2 Loran Overview	269
11.2.1 Loran-C	269
11.2.2 eLoran	271
11.3 Theory of Operation	272
11.4 Historical Reasons for GNSS/Loran Integration	275
11.5 Integration Scenarios	276
11.5.1 Position-Domain Integration	276
11.5.2 Range-Domain Integration	278
11.5.3 Déjà Vu Navigation: A Case Study of Range-Domain Integration	281
11.5.4 Integrity with Range-Domain Integration	283
11.5.5 Improved Accuracy for Loran Integrity	286
11.5.6 Tracking Loop Domain Integration	287
11.6 Conclusions	288
References	288

CHAPTER 12

Indoor and Weak Signal Navigation	291
12.1 Introduction	291
12.2 Signal Processing Considerations Related to Weak Signals	292
12.2.1 Acquisition of Weak Signals	294
12.2.2 Clock Stability and Integration Times	295
12.2.3 Tracking of Weak Signals	296
12.2.4 Cross-Correlation and Interfering Signals	297
12.2.5 Multipath Mitigation	298
12.2.6 Benefits of Future GNSS	300
12.3 Aiding Possibilities and Supportive Systems	300
12.3.1 Assistance	300
12.3.2 Supportive Systems for GNSS	301
12.4 Navigation Algorithms for Difficult Signal Conditions	303
12.4.1 Constraints on User Motion	304
12.4.2 Map Matching	305
12.4.3 Adaptive Algorithms	305
12.5 Quality and Integrity Monitoring	306
12.5.1 Introduction to Integrity Monitoring	306
12.5.2 Reliability Testing	307
12.5.3 Weighted Least-Squares Notation	308
12.5.4 Residuals and Redundancy	310
12.5.5 Global Test	311

12.5.6	Local Test	312
12.5.7	Null Hypothesis and Alternative Hypothesis	314
12.5.8	Parameters for Fault Detection and Exclusion	314
12.5.9	Multiple Outliers	315
12.5.10	Fault Detection and Exclusion in Kalman Filtering	316
12.5.11	Quality Control	316
12.5.12	The Practical Side of Quality Control	317
12.6	Examples Included on the DVD	319
12.6.1	Example 1: Acquisition of Weak Signals	319
12.6.2	Example 2: Fault Detection and Exclusion	322
12.7	Summary	323
12.8	Further Reading	324
	References	324

CHAPTER 13

Space Applications	329
13.1 Introduction	329
13.2 Operational Considerations	329
13.2.1 Spacecraft Velocity	330
13.2.2 Orbit Geometry	330
13.2.3 Antenna Direction	332
13.2.4 Size and Power	332
13.2.5 Multipath	333
13.2.6 Signal Strength	333
13.2.7 Environment	334
13.3 Applications	335
13.3.1 Precise Orbit Determination	335
13.3.2 Real-Time Navigation	335
13.3.3 Formation Flying and Proximity Operations	336
13.3.4 Remote Sensing	337
13.3.5 Attitude Determination	338
13.3.6 High-Altitude GNSS	339
13.3.7 Launch, Entry, and Landing	340
13.4 GNSS Modernization	340
13.5 Example: Processing Raw Measurements from the GRACE Satellite	341
13.6 Summary	344
References	344

CHAPTER 14

Geodesy and Surveying	347
14.1 Introduction and Background	347
14.1.1 GNSS Surveying	348
14.1.2 GNSS Geodesy	349
14.2 Technical Overview	350
14.2.1 The Data Models and Processing Strategies of GNSS Geodesy and Surveying	350

14.2.2 Mathematical Models	351
14.2.3 Baseline Processing	356
14.2.4 Network Processing for Positioning	359
14.3 GNSS Ground Infrastructure—Continuously Operating Reference Station (CORS) Networks	360
14.3.1 The IGS Infrastructure	361
14.3.2 National CORS Infrastructure	364
14.4 Surveying and Geodesy Applications and Operational Modes	367
14.4.1 GNSS Surveying	368
14.4.2 GNSS Geodesy	372
14.5 The Future: The Next-Generation GNSS	376
14.5.1 The Benefits of More Satellites and Signals	376
14.5.2 Improvements to the GNSS Infrastructure	377
14.5.3 Applications and the Future	378
References	379

CHAPTER 15

Atmospheric Sensing Using GNSS Occultations	381
15.1 Introduction	381
15.2 Occultation Measurements	382
15.3 Atmospheric Retrievals	384
15.3.1 Derivation of Bending Angle Profiles	385
15.3.2 Ionospheric Calibration	387
15.3.3 Derivation of Atmospheric Profiles	388
15.4 Weather and Climate Applications	390
15.5 Recent Advances	392
15.6 Scripts and Data Included on the DVD	394
15.7 Further Reading	394
References	395

CHAPTER 16

Remote Sensing Using Bistatic GNSS Reflections	399
16.1 Introduction	399
16.1.1 General Discussion of Traditional Remote Sensing	400
16.1.2 Remote Sensing Using Reflected GNSS Signals	401
16.2 Reflection Geometry	402
16.2.1 Estimating the Surface Reflection Point Location	403
16.2.2 Delay and Doppler Spreading over the Surface	403
16.3 Signal Processing	403
16.3.1 Detection and Surface Mapping	405
16.3.2 Averaging Consecutive Correlations	407
16.3.3 Delay Waveforms and Delay Doppler Maps	408
16.4 Remote Sensing Theory	410
16.4.1 Bistatic Surface Scattering	410
16.4.2 The Bistatic Radar Cross Section	413
16.4.3 Sea Surface Modeling	414

16.4.4	Bistatic Scattering from Land	416
16.4.5	Bistatic Scattering from Sea Ice	417
16.5	Ocean Altimetry	418
16.5.1	Motivation	418
16.5.2	Aircraft Altimetry Measurements	418
16.5.3	GNSS Ocean Altimetry from Space	420
16.6	Ocean Wind and Wave Sensing	421
16.6.1	Aircraft Wind and Wave Measurements	421
16.6.2	Wave Sensing from Spacecraft	423
16.7	GNSS Bistatic Land and Ice Sensing	424
16.7.1	The History and Applications of GNSS Land Reflections	424
16.7.2	Spacecraft-Detected Land Reflections	426
16.7.3	The History and Applications of GNSS Ice Reflections	427
16.7.4	Spacecraft-Detected Sea Ice Reflections	428
16.8	Data Provided on the DVD	432
16.8.1	Specular Point Calculation Scripts	432
16.8.2	Surface Scattering Model	433
16.8.3	Spacecraft Data and Processing Tools	433
16.9	Further Reading	433
	References	434

CHAPTER 17

	New Navigation Signals and Future Systems in Evolution	437
17.1	The History of GNSS	437
17.1.1	GPS	437
17.1.2	Modulation of Satellite Carrier Signals	439
17.2	Motivation for Evolution	439
17.2.1	Main Concept of Operation for Galileo	440
17.3	New Modulation Opportunities	441
17.3.1	Existing Spreading Symbol—BPSK Modulation	442
17.3.2	Binary Offset Carrier (BOC) Modulation	445
17.3.3	Multiplex BOC Modulation	453
17.3.4	Composite BOC Modulation	455
17.3.5	Time Multiplex BOC Modulation	457
17.3.6	Other Spreading Symbol Modulation Options	459
17.3.7	Alternative BOC (AltBOC) Modulation	461
17.4	Signal Multiplex Techniques	465
17.4.1	QPSK	466
17.4.2	Interplex	467
17.4.3	Other Techniques	468
17.5	Interference	468
17.5.1	Performance Metrics	470
17.5.2	Spectral Separation Coefficients (SSC)	473
17.6	Listing of Proposed Systems and Signal Characteristics	478
17.6.1	Global CDMA Satellite Navigation Systems I: GPS	479
17.6.2	Global CDMA Satellite Navigation Systems II: Galileo	480
17.6.3	Global CDMA Satellite Navigation Systems III: COMPASS	481

17.7 Summary	482
References	483
About the Editors	485
About the Contributors	485
Index	491

Preface

Over the last decade, the number of applications that utilize global satellite navigation systems (GNSS) has steadily increased. Perhaps the most visible manifestation of this change is the development and wide use of consumer technologies that are based on GNSS. For example, these days it is not unusual for automobiles or mobile phones to be equipped with a Global Positioning System (GPS) navigation system. Additionally, scores of applications have emerged that even the original system designers could not have imagined. One such example is environmental remote sensing. In view of this trend, it is reasonable to expect that more satellite navigation based applications, in parallel with new satellite navigation systems, will be arriving in the future.

In the past, the group of professionals that were familiar with the detailed workings of GNSS came from a select group. As the number of applications grows, so will the need for professionals to understand this technology and how to apply it. As new applications emerge and existing applications expand, this group of knowledgeable professionals is growing to include individuals in a wide range of technical disciplines. The primary motivation for writing this book was to provide what we believe is an important need for this widening field of professionals—to provide the professional (and, perhaps, even hobbyist) interested in GNSS the opportunity to gain a practical or *hands-on* experience working with algorithms, applications, and data. We hope this book will allow the reader to learn about GNSS by “doing,” as we believe it is the most effective way to understand a complicated subject. This book will allow readers to expand on the basic theory behind GNSS systems and apply new insight to practical engineering and scientific applications.

A second motivation for this book is the fact that until now, the books most individuals relied on to learn about basic GPS applications have been the seminal two-volume set *Global Positioning System: Theory and Applications* edited by B. Parkinson, J. Spilker, P. Axelrad, and P. Enge. These books, informally known as the “GPS blue books,” have been a valuable resource for nearly every student of satellite navigation and will remain a useful reference for years to come. They could very well be on a shelf within arms reach of many readers. However, since the initial publication of the “GPS blue books,” several advancements have occurred in various traditional fields of navigation and GNSS. For example, microelectromechanical systems (MEMS)-based inertial sensors as well as inexpensive, compact, and powerful computing resources have become readily available. The availability of computer processing tools and the relative ease with which some of these applications can be both described and *demonstrated*, convinced us that it is possible to

provide a more hands-on approach for those wishing to learn about GNSS and GNSS applications. It is now possible to give interested professionals the chance to “have a go” at implementing GNSS-based algorithms and applications themselves. We believe that providing an additional resource that delves into as many new applications as possible, accompanied by numerous practical examples, all compatible with freely available tools and released under free and open source licenses, would be a welcome new resource for developers.

This book is not intended to provide a comprehensive treatment of GPS or satellite navigation in general terms. Nor is it intended to replace the “blue books” or even some currently available excellent texts such as *Global Positioning System: Signals, Measurements and Performance* (Second Edition) by Pratap Misra and Per Enge, or *Understanding GPS: Principals and Applications* (Second Edition) edited by Elliot Kaplan and Christopher Hegarty. Instead, our intent is to provide a book that compliments these fundamental texts, providing an additional resource for scientists and engineers who are looking for a basic overview of GNSS principals and practical introductions to a wide range of navigation applications.

This book will allow individuals to start “digging” deeper into specific applications where GNSS plays an important role and start answering the question: How is GNSS utilized to solve this particular problem? For those already familiar with the operation of GNSS receivers but interested in pursuing a specific application, this book will give them useful examples. For the adventurous, it will allow them to explore new areas and (hopefully) encourage them to ask “Where do I go from here?” or “What’s next?”

To this end, the book is organized as follows: Chapters 1 through 5 present the basics of GNSS navigation, or what could be considered traditional uses of GNSS. This introduction to basic GNSS principles includes: GNSS signal acquisition and tracking; GNSS position, velocity, and time (PVT) estimation; and differential GPS techniques and algorithms. These chapters include both a GNSS measurement simulator to demonstrate traditional position and velocity applications, and a complete software implementation of a GPS receiver, which processes actual off-air signals.

Following are several chapters that address GNSS integrated with other navigation systems. These include GNSS combined inertial navigation systems (Chapters 6 and 7); LADAR (Chapter 8); various radio frequency (RF) measurements (Chapter 9); and terrestrial radio navigation systems such as LORAN (Chapter 11).

Additionally, there are numerous chapters that address specific application areas or unique uses of GNSS signals. These chapters cover such broad application areas as: aviation (Chapter 10); indoor and weak signal navigation (Chapter 12); spacecraft applications (Chapter 13); and geodesy and surveying (Chapter 14). Finally, two novel scientific applications will be presented demonstrating how GNSS signals can be used for remote sensing of the environment (Chapters 15 and 16). We hope these chapters will expand the reader’s horizon about the possible uses of GNSS.

Of course, the only GNSS that has consistently been available and operational over the last decade is GPS. In anticipation of GPS modernization efforts, an overview of new signals and systems being developed for GPS is provided. A related topic to this is the emergence of new GNSS constellations such as the European

Galileo project. Both of these topics are the subject of the final chapter of the book (Chapter 17). For more details on the specific content of each chapter and the tools and data sets available on the DVD included with this book, please refer to the summary road-map included in Chapter 1.

In closing, we would like to acknowledge the fact that we had considerable support from many individuals in preparing this book. While it would be impractical to list all individuals that helped in preparation of this book, we would like to acknowledge those individuals that helped review various revisions of the manuscript and provide critical feedback, and those that tested and evaluated the software routines that go along with the various chapters in the book. In alphabetic order of last names these individuals are Mounir Adjrad, Vibhor Bageshwar., Susmita Bhattacharyya, Paw Yew Chai, Maria Paola Clarizia, Jason Da Ponte, David De Lorenzo, Paul Groves, Steve Hilla, Thomas Jakel, Alexander Mittleman, Oliver Montenbruck, Curtis Olson, Rohit Pandita, Todd Walter, Gary Wick, Zhiqiang Xing, and Guijin Zheng.

Finally, we would like to note that all of the editor and contributor royalties from the sale of this book will be split between the following charitable organizations: Médecins Sans Frontières (Doctors Without Borders) and The Institute of Navigation's student scholarship and awards program.

For additional information on many of the applications and code included within this book, readers are encouraged to visit the Web site, www.gnssapplications.org.

Scott Gleason
Demos Gebre-Egziabher
August 2009

Global Navigation Satellite Systems: Present and Future

Scott Gleason and Demoz Gebre-Egziabher

1.1 Introduction

While there is no universally agreed upon definition of the term navigation, this book uses it to describe the process of determining the position, velocity and, in some instances, the attitude (or orientation) of an object. The object (a vehicle, for example) whose position, velocity, and attitude we are interested in knowing is called the user, and the system that generates the position velocity and attitude solution is the navigator. Together the position, velocity, and attitude solution are called the navigation state vector of the user. An accurate and reliable means for generating a user's navigation state vector is indispensable for many applications. However, until the advent of global navigation satellite systems (GNSS), precise navigation was something reserved for the very skilled navigator or those who could afford to purchase an expensive navigation system. Today, GNSS has made precise navigation affordable for the masses. They are used to generate an estimate of position, velocity, and time by processing signals transmitted from satellites in known orbits. The single most important and ubiquitous utility in modern navigation today is the GNSS known as the global positioning system (GPS). Its accuracy, universal availability, and low cost have made it the navigation and timing system of choice for many users.

Over the past decade the number of military, commercial, and scientific applications that leverage accurate positioning and timing information has increased dramatically. Interestingly, many of these applications were not or could not have been envisioned by the original system designers. For example, many of these applications aim at using GPS in environments in which it was not originally intended to be used. That is, GPS was primarily designed for outdoor applications with clear views of the sky. This expansion into new user environments has resulted in numerous new applications of GNSS, many beyond navigation and guidance. The current popularity of many of these applications makes it reasonable to expect that the number of new and novel GNSS-based applications will continue to increase.

This book was written to provide readers with a fundamental knowledge of GNSS the opportunity to develop a deeper working knowledge of these systems. This can be done quickly through a hands-on experience with the algorithms used to realize many of the fundamental GNSS applications. This book also introduces readers to the many novel applications of GNSS, including uses of the system

beyond the basic navigation and timing applications for which it was originally designed. To this end, this chapter starts with an overview of existing GNSS constellations. Next, it discusses the many popular applications that have resulted from the use of GNSS. Following is a discussion of the features and issues associated with the planned modernizations of existing GNSS constellations and new ones under development. This will include a brief discussion of modern GNSS receiver technology. Finally, the chapter concludes with a roadmap of this book and a suggested approach for using the material provided.

1.1.1 Current and Planned GNSS Constellations

GNSS are constellations of satellites designed to provide positioning and timing information for users on Earth or in space. Currently, the most widely used operational GNSS is the GPS. It consists of nominally 24 satellites in approximately 12-hour orbits (even though currently there are more than 24 in orbit). The system was designed and realized by the U.S. Department of Defense, and since being commissioned, it has expanded in ways that few would have predicted. In addition to the military applications, a large suite of commercial and public sector users and applications have appeared.

A representative but nonexhaustive list of such applications includes: placing GPS receivers on ocean buoys to measure wave height and direction [1], monitoring the Earth's crustal deformations [2], and sensing the atmosphere using occultation techniques, such as those accomplished using the CHAMP satellite [3].

In addition to GPS, another GNSS that was fully operational in the recent past was the Russian Federation's GLONASS. GLONASS maintained a full constellation of 24 satellites (briefly in 1996), which was reduced to eight to ten working satellites in 2001 [4]. In recent years, uncertainty about the future of the GLONASS system has limited the demand and availability of receivers to process the signals. While it is not clear what the future of GLONASS will be, currently, it appears that there is a resurgence in the availability of GLONASS signals and receivers that can process these signals.

Despite the popularity of GPS, many users are interested in alternative systems. This is motivated, in part, by the fact that GPS is a system operated and controlled by the U.S. Department of Defense. As such, some users may want to retain a navigation capability that is not solely GPS-based. Another, more technical, motivation for this is due to the fact that GPS (or any other GNSS) is a single system, and it is conceivable that a single failure can result in a denial of service to a large number of users. Multiple GNSS may provide a level of redundancy and, thus, an added degree of robustness to GNSS applications.

The Galileo constellation is the European Union's effort to provide both an alternative and a compliment to GPS. The first Galileo satellite was launched in December 2005 and transmitted test signals shortly afterward [5]. The Galileo constellation is tentatively expected to become operational sometime after 2010.

In addition to GPS, GLONASS, and Galileo there are other GNSS currently in development, such as the China's COMPASS constellation [6]. This GNSS is being designed and implemented by China and currently has a growing user base in that region. COMPASS is currently in its in-orbit validation (IOV) phase, with

plans to be operational in the Asia-Pacific region by 2010. Numerous additional launches are scheduled in the coming years with a planned fully operational constellation (FOC) of 5 geostationary and up to 30 medium Earth orbit satellites providing global coverage. Chapter 17 provides a more detailed discussion of future GNSS and their signals.

1.1.2 GNSS User Architectures

It is possible to navigate with GNSS signals using a variety of configurations. The following sections provide an overview of the most common setups.

1.1.2.1 Stand-Alone Satellite Navigation

This is the basic method of GNSS navigation where only the received signals from a GNSS constellation, such as the publicly available GPS standard positioning service (SPS) are used. This includes applications such as assisting boats to find their way in and out of harbors using only a stand-alone receiver. As will be demonstrated in Chapters 3 and 4, the performance of stand-alone GNSS is sufficient only for a limited number of applications. Many applications either desire or require higher accuracy than a stand-alone SPS can provide. For this reason, GNSS is often combined with other sensors and signals.

1.1.2.2 Differential GNSS (DGNSS) Navigation

Differential systems are primarily intended to improve the stand-alone accuracy of a GNSS receiver position estimate, while also providing information on the position integrity. Surveying engineering work is an example of an application that often uses differential carrier phase GNSS. Figure 1.1 shows the basic configuration of an example DGNSS implementation.

“Differential” indicates that a difference is being taken to mitigate some of the errors present in the stand-alone navigation estimate in an attempt to improve users’ knowledge of their position. This will typically consist of a reference system measuring some of the satellite system errors and relaying this information over a radio frequency link to users in the vicinity. An example of such a system providing coverage in North America is the wide area augmentation system (WAAS).

1.1.2.3 Network-Assisted GNSS (A-GNSS) Navigation

Whenever a communications network is used to relay information to a GNSS receiver, it can be said to be receiving assistance. This is called assisted GNSS or A-GNSS. In this regard, DGNSS described above can be thought of as a subset of A-GNSS. This assistance is often a correction to raw measurements calculated elsewhere and sent over a radio link to remote receivers. However, unlike DGNSS, in A-GNSS this assistance can often include more basic information used to assist the receiver in performing an accelerated position fix or to extend the validity of the satellite information used during positioning. For example, each GNSS satellite

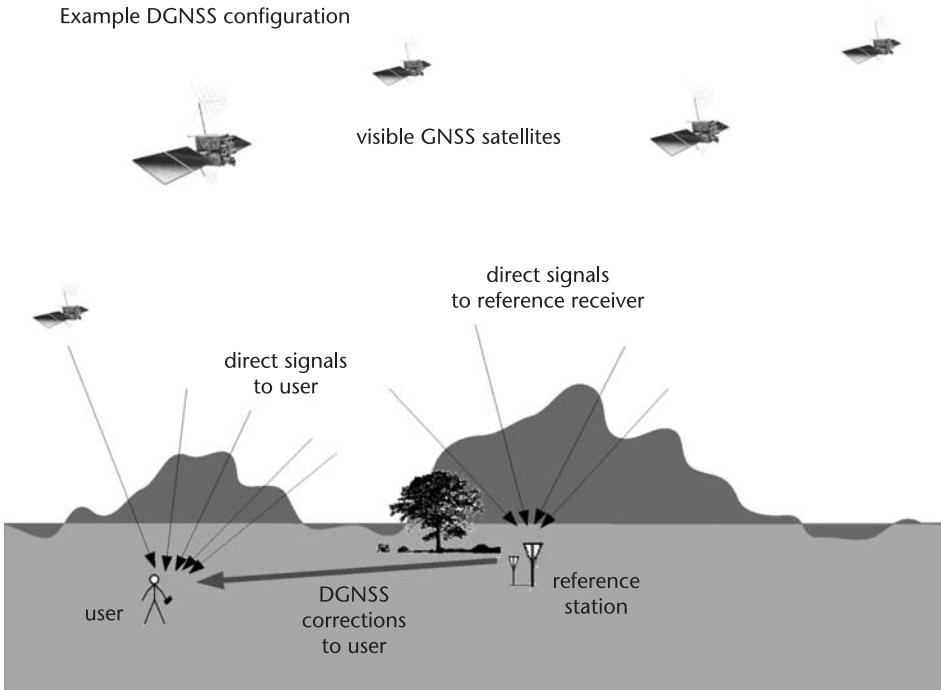


Figure 1.1 A typical example of a DGNSS user configuration.

transmits data on its signal with detailed information on its orbit, clock parameters, integrity, and status information. This information is required to estimate the user position. The relatively slow rate (i.e., 50 Hz or 20 ms per data bit in the case of the GPS L1 C/A signal) at which this information is modulated onto the satellite signal is seen as a drawback. However, in an A-GNSS, the necessary satellite information as well as the approximate location of the user can be determined externally (using reference stations and a central server for example) and sent to the user receiver upon request to help provide a much faster time to first position fix. This is a method employed widely in the implementation of mobile phone-based and in-car-based satellite navigation systems. Figure 1.2 illustrates an A-GNSS configuration.

Using satellite navigation data in this way holds the potential for greatly improving GNSS receiver performance and reducing the processing load required in the receiver. As will be discussed in more detail later, the advantage of A-GNSS is particularly relevant in the area of software-based GNSS receivers where the management of the processing load is of great importance (see Chapter 5).

Additionally, it is worth noting that the navigation data contained on the satellite signal is only valid for a short period of time (conservatively, only a few hours). This leaves open the opportunity—which many have jumped in to fill—to *extend* the usable duration of the satellite orbit parameters. These methods use modeling techniques to predict the locations of the GNSS satellites with sufficient accuracy for positioning up to several weeks into the future. In a network-assisted

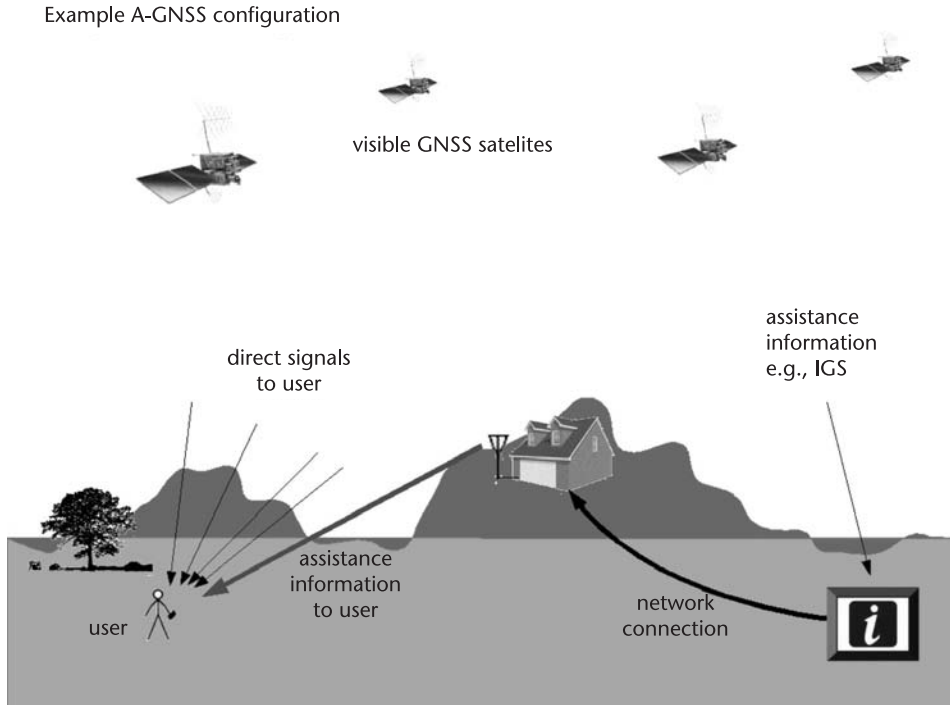


Figure 1.2 A typical example of an A-GNSS user configuration.

system, such an extended period of validity is very useful, especially if the receiver is traveling in a difficult environment where tracking gaps resulting in incomplete navigation data messages from the satellites are common.

1.1.2.4 GNSS and Integrated Navigation

While GNSS provides an unprecedented level of accuracy and ubiquity of navigation, it has well-known shortcomings. Many of these disadvantages can be significantly curtailed by using integrated navigation systems. For example, one such shortcoming is the signal's susceptibility to unintentional or malicious radio frequency interference (RFI) or jamming, and the fact that the signal cannot provide an attitude or orientation solution easily—a feature that is indispensable in many vehicle navigation and guidance applications. Even though the use of GPS in attitude determination has been demonstrated with considerable success, it requires specialized receivers and multiple antennas separated by some distance. Both the above noted drawbacks can be mitigated by the use of multisensor or integrated navigation systems. Multisensor or integrated navigation systems combine the information from multiple sensors to generate an estimate of a vehicle's navigation state vector.

Inertial navigation systems (INS) have been integrated with GNSS with considerable success. This fusion between GNSS and INSs is complementary: The INS helps mitigate the shortcoming of the GNSS and vice versa. Other sensors that are

also commonly integrated with GNSS include but are not limited to baro-altimeters, LIDAR, magnetometers, and other radio frequency (RF) signals.

1.1.2.5 GNSS Indoor Navigation

Satellite navigation was primarily designed for outdoor applications with line of sight visibility to the satellites. However, there is an increasing demand for a navigation device that functions reliably indoors. Chapter 12 will describe in more detail how weak and reflected signals, like those often found indoors, can be used to generate a position estimate. As mentioned above, it is often the case that GNSS signals need to be augmented with external sensors to function accurately indoors. These sensors might include other RF signals at frequencies that penetrate buildings, or other external sources of information that can be used when GNSS signals are not available. A system for reliable indoor navigation using GNSS is still the subject of much research and will be helped by the planned addition of new signals and satellites.

1.1.2.6 Location-Based Services

A rapidly growing use of satellite navigation is in applications known as location-based services. This refers to the use of position information to augment or enable additional features on existing devices, such as mobile phones, and can be applied in conjunction with all of the user architectures mentioned above. In these applications the main concern is not the position of the user, but how this position can be linked to the user's environment. A common example of this is someone searching for a restaurant or pub close to where they are located. In this case, the position of the user, as provided by a GNSS or otherwise, can be used as a reference to retrieve a map that highlights points of interest in the user's vicinity. This integration of GNSS position information with external maps and databases is an important trend in numerous applications, as summarized in Section 1.1.3.

1.1.3 Current GNSS Applications

Currently GPS is the GNSS primarily used by both military and civilian users. However, future users will have several additional GNSS options at their disposal as new systems come online. These GNSS signals, which, in many cases, are free and globally available, will be used to advance applications that were pioneered using GPS. Regardless of the GNSS used, in this book we group applications of GNSS into the following categories:

- *Personal navigation:* This consists of applications to aid people on foot to navigate. An example of this is widely known personal navigation devices used by hikers. This technology is currently making its way into cell phones and PDAs, expanding satellite navigation to a large new group of users. Chapters 2–5 cover several of the standard techniques used in basic navigation.

- *Aviation applications:* This includes en route navigation as well as the precision approach and landing applications, which demand a very high level of performance in terms of accuracy and robustness. Chapter 10 covers this topic in more detail.
- *Automotive applications:* GNSS-based systems have enabled the development of many automotive applications that enhance ease of operation. The simplest and most widely used systems provide drivers with instructions to get from departure point to destination with a minimum of detours. There are, however, more sophisticated GNSS-based applications that can enhance the safety of automotive operations. One example application is that of estimating vehicle parameters in real time to improve vehicle handling characteristics. Chapter 7 gives an example of one such application.
- *Weak signal navigation:* These are applications where the GNSS signal quality is poor. For example, in certain applications such as indoor navigation, the GNSS solution has to be optimized for robust performance in environments where the signal is highly attenuated. In these applications stand-alone GNSS positioning is not reliable. Chapter 12 covers the topic of weak signal navigation in detail.
- *Marine applications:* This is one of the original civil applications for GPS, and GPS is well-suited to this use. This is because of the clear views of the sky and modest accuracy requirements of most maritime applications. GPS receivers have become standard equipment on boats of all sizes today, and they perform a very valuable service to the global maritime community. Most marine applications use some form of stand-alone navigation, which is described in Chapter 3.
- *Space applications:* GPS receivers have proven to be a very valuable tool on board Earth-orbiting satellites. They have been principally used in low Earth orbit (LEO) satellites, but their use is currently expanding into space vehicles operating at higher altitudes. In space applications, GPS has the potential to be used as an attitude determination sensor. Chapter 13 covers this topic in more detail.
- *Agriculture, forestry, and natural resource exploration:* These diverse applications include geological monitoring, forest management, mining, and oil exploration. These applications often combine GNSS field measurements with geographic information system tools to produce accurate regional maps for resource monitoring and management. There are several existing books on this subject that the interested reader is encouraged to consult.
- *Geodesy and surveying:* This application is perhaps the best example of the direct benefits that have resulted from the public availability of GPS signals. Geodesy applications require precision positioning information at the centimeter or millimeter level and include applications such as the monitoring of the movements of the Earth's crustal plates or ice shelves, often involving extensive postprocessing. Similarly, GNSS surveying has become widespread and generally has more relaxed accuracy requirements. These two applications are covered in Chapter 14.
- *Scientific applications:* In addition to geodesy and surveying there have been attempts to use GPS in a number of scientific research fields. These include

using GPS signals for remote sensing of the environment and space weather studies, as described in Chapters 15 and 16.

While the above is by no means an exhaustive list of applications, it shows the utility and versatility of GNSS. For example, although not covered in this book, a large and rapidly expanding application area is that of timing and frequency standards. In this application, GNSS-based time is used to synchronize large telecommunications networks. The utility and versatility increases more when we consider the fusion of GNSS with other complementary sensors. This enables applications that would not be possible with GNSS alone. This increase in performance can occur in one of the following three ways.

1. The other complementary sensors can provide a position, velocity, or timing solution when GNSS signals are temporarily unavailable or are of poor quality.
2. The information from the other complementary sensors enhances the robustness and accuracy of the GNSS receiver's ability to track the signals. That is, the information from the other sensors can be used to affect the way in which a GNSS receiver processes the received satellite signals.
3. Acting in concert, the complementary sensor and GNSS allow applications that neither sensor alone can provide. Chapters 6–9 discuss the integration of GNSS with other sensors. Examples of other sensors that can be combined with GNSS are inertial measurement units (Chapters 6 and 7), LADAR (Chapter 8), other radio frequency transmitters (Chapter 9), and Loran (Chapter 11).

1.1.4 Positioning Performance Measures

Even though one normally hears of GNSS performance being measured by a single scalar quantity such as *accuracy*, a complete characterization of its performance requires using multiple metrics. This is because the quality of performance sought by different users is different and reflects their particular application. For example, aviation users of GNSS demand that the navigation solution they use minimizes their *integrity* risk. Integrity is the measure of the likelihood of occurrence of an undetected navigation error that can be hazardous. In addition to accuracy and integrity, two other metrics that are normally used to characterize a GNSS's performance are *continuity risk* and *availability*. Continuity risk is the probability of a detected but unscheduled navigation function interruption after an operation has been initiated. Availability is defined (or computed) as the fraction of time a navigation system is providing a solution to the specified level of accuracy, integrity, and continuity. From these definitions, it is clear that the performance of a navigation system depends on the environment and particular scenario it is being used for at a given instant. Thus, it is difficult to provide meaningful statements quantifying the performance of a given GNSS-based system without taking into account its application. A more meaningful simple comparison, therefore, is to *qualitatively* describe how a given GNSS augmentation system performs relative to a stand-alone GNSS. Table 1.1 provides such a comparison.

Table 1.1 Qualitative Comparison of Performance Enhancement Afforded by Various GNSS User Architectures

	<i>Accuracy</i>	<i>Availability</i>	<i>Integrity</i>	<i>Continuity</i>	<i>Coverage (Service Volume)</i>
Stand-Alone	—	—	—	—	—
DGNSS	Increased	Increased	Increased	—	Same
A-GNSS	Same	Increased	N/A	Increased	Increased
GNSS/INS	Same/Increased	Increased	Same/Increased	Increased	Same/Increased

1.2 GNSS Signal Improvements

As the GPS user base has increased, so has the demand for higher accuracy and robustness. In its current configuration, GPS alone cannot provide the demanded level of performance for many applications, particularly those demanding a high level of integrity (robustness). This, in part, has been the impetus for planned future upgrades to the GPS constellation and the advanced features being included in the design of other GNSS, such as Galileo.

One method for improving the accuracy of a GNSS is to add new signals that, by themselves or used together with the existing signals, provide inherently better error performance. In the case of GPS, the U.S. government announced in 1999 that it would add signals for use alongside the existing signals [7]. The new civil (i.e., publicly accessible) signals will include an additional signal on the existing L2 frequency, which will be known as L2C. Additionally, a new civil signal broadcast at 1.17645 GHz and called L5, will be included on all future IIF satellites. There are also plans for two new military signals on the existing L1 (1.57542-GHz) and L2 (1.2276-GHz) frequencies, known as M code signals. Figure 1.3 [8] illustrates the existing and new GPS signals.

Several of the coming improvements are described below, with more emphasis placed on the signals available to the general public or civil users. For more detail,

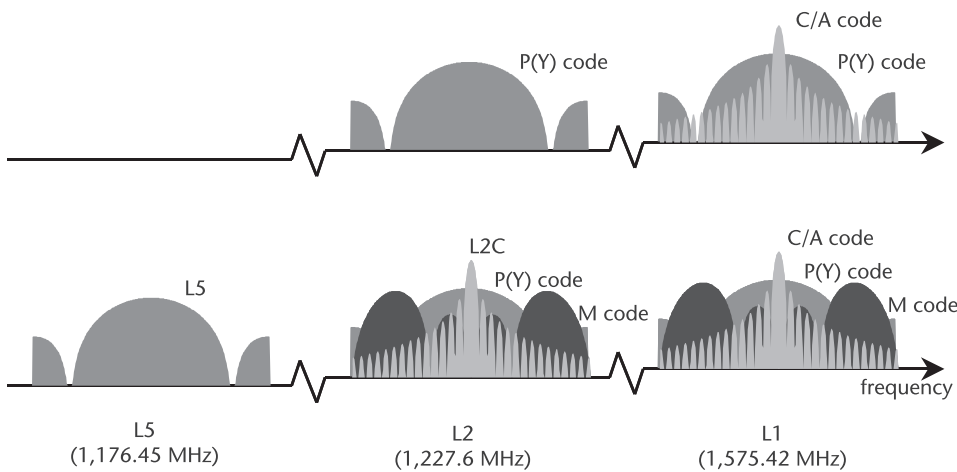


Figure 1.3 Existing and new signals. (Source: [8]. Reprinted with permission.)

readers should consult the GPS or Galileo interface control documents included on the DVD, such as ICD-200D [9].

1.2.1 Additional GPS Frequencies

Currently, GPSs transmit only one publicly accessible signal—the clear acquisition (C/A) code on the L-band carrier known as L1. Until May 2000, intentional degradation of the signal's quality via selective availability (SA) was the primary limitation to accuracy. Since the discontinuation of SA, currently the largest and most persistently troublesome of these errors are those caused by the unknown delays and other disturbances added by the Earth's atmosphere. One way to reduce these errors is by using differential corrections, which requires external (i.e., beyond the receiver) infrastructure. Alternatively, the errors induced by the ionosphere can be significantly reduced when measurements on two different carrier frequencies are used. This results in a significant improvement of the positioning accuracy. This is one of the reasons that the GPS constellation has included plans for additional civilian signals on the L2 and L5 frequencies. These two new signals on future GPS satellites will allow civil users to greatly improve their stand-alone positioning accuracy by significantly reducing the errors accumulated as the signals travel through the Earth's atmosphere.

The Galileo system is being designed with this in mind and like GPS will include publicly available signals at three different L-band frequencies currently designated as E1 (overlapping GPS L1), E5 (1.191715 GHz), and E6 (1.278750 GHz) [10].

An additional benefit of using measurements at several frequencies concurrently is the possibility of using a technique known as *widelaning*, which enhances the robustness of carrier phase-based positioning. Currently, single-frequency carrier cycle ambiguity resolution and *verification* techniques require relatively long and continuous signal tracking and processing, negatively impacting the robustness of the carrier-based positioning solution. The use of multiple frequencies would greatly reduce the time required for carrier-phase ambiguity resolution and verification. More information on widelaning techniques and their advantages can be found in [11].

1.2.2 Higher Accuracy Ranging

Range estimates are obtained in a GNSS receiver by tracking the unique time-based signatures of the satellite transmissions. At any given instant if we visualize a GNSS signal as a large ruler extending from the transmitting satellite to the user receiver, then these unique time-based signatures correspond to the marking (or ticks) on the ruler. The accuracy of a range estimate from a GNSS satellite will depend on how accurately the markings on the ruler can be observed and distinguished. For the case of the GPS C/A code signal, the coarse markings are spaced at roughly 300-m intervals. Within or between the coarse markings, ranges can be estimated to about 1m. Improvements can be made to this estimation process if the time characteristics of the signals are at shorter intervals. Continuing with the large ruler analogy, this means it is necessary to make the spacing between the coarse ruler markers smaller. This can be done by designing a tracking code with chip

widths less than 300m. With higher resolution ranging information (obtained by modulating the ranging code onto the signal at a higher frequency) the precise distance between the transmitter and receiver can be measured more accurately, thereby improving the receiver position estimation.

In the case of the new GPS L5 signal, the ranging code frequency has been increased by a factor of 10, where a single chip will be approximately 30m in length, resulting in significant improvement in the accuracy of range estimates. Additionally, the Galileo E5 and E6 signal timing codes are all transmitted at higher rates with this improvement in mind. A shorter code length will also present new challenges, as features often taken for granted using the GPS L1 signal, such as rapid signal acquisition, may become more challenging.

1.2.3 Longer Ranging Codes

The current GPS L1 C/A ranging codes repeat every millisecond. Under most circumstances this does not result in any problems, for the codes were designed as unique sequences, easily distinguishable from one satellite to another. In practice, however, especially under weak signal conditions, it is often possible to observe what are called cross-correlations between different satellite transmissions. In other words, a receiver may start to see traces of the ranging code from satellite Y while looking to make measurements from satellite X. This can cause confusion within the receiver and lead to errors in the range measurements.

To address this in the next generation of GNSS signals the individual ranging codes for each satellite should be extended well beyond 1 millisecond in length, often using what are called secondary codes. By extending the length of the ranging code it is possible to eliminate these pesky cross-correlations during signal processing. This acts to increase the reliability with which the signal from each satellite can be tracked. For an example of how cross-correlations can cause problems in weak signal conditions, we recommend experimenting with the hands-on acquisition examples in Chapter 12.

The next generation of GPS signals at L2 and L5 will include ranging codes of increased lengths. In the case of L2 this increase will be to 20 and 1,500 milliseconds [8]. In the case of Galileo, all the transmitted signals (E1, E5, and E6) are modulated with ranging codes of lengths significantly greater than 1 millisecond [10].

1.2.4 Higher Transmit Power Levels

Ideally a GNSS receiver should work at all times and in any environment. For example, hikers walking around the forest in areas with a dense tree canopy or callers using their GNSS-equipped cell phones in an office building should still be able to estimate their location. As it turns out, often this is not the case due to the limited capability of GNSS signals to penetrate into environments, such as dense forests and office buildings. So, in addition to improved processing methods, the ability to track the next generation of GNSS signals will be enhanced, because the signals will be transmitted from the satellites at higher power levels. This should

help the signals to penetrate into areas where current signals are very difficult to detect and use reliably.

1.3 Advanced Receiver Technology

A GNSS receiver is a combination of hardware and software capable of receiving signals from several GNSS satellites and processing them into useful position, velocity, and timing information. Figure 1.4 shows a general block diagram of the components required in a GNSS receiver. Many advances in GNSS receiver design are being considered to enhance overall navigation, guidance, and timing functions. To put these advances in context, however, we need to briefly describe current conventional GNSS receivers.

1.3.1 Conventional Receivers

A conventional GNSS receiver architecture consists of an antenna linked to a series of application-specific integrated circuits (ASICs), all controlled by a central processor. Working together, this combination of hardware and software performs all of the receiver functions. A common configuration example consists of an antenna and an RF front-end chip [12], which then feeds the sampled signals into digital correlation chips [12]. All of the subsequent signal processing steps are controlled using interfaces from the ASIC to a central processor. The individual blocks in Figure 1.4 are described in Chapter 5 in the context of a software-only based receiver. The end result is an estimation of the position, velocity, and time (PVT) of the receiver.

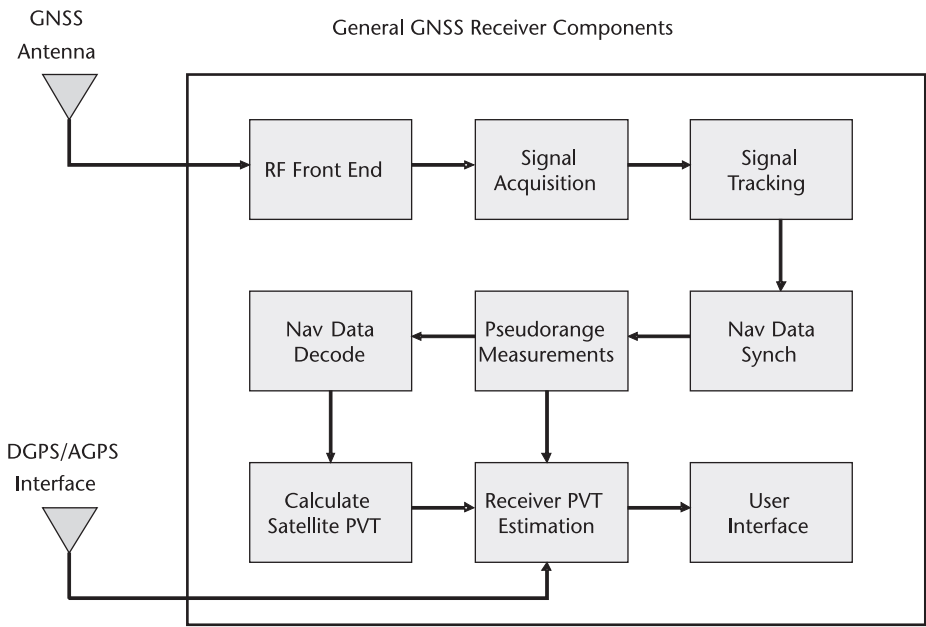


Figure 1.4 Block diagram of a typical GNSS receiver.

In this case, the bulk of the receiver algorithms are executed in the main processor using register or memory interfaces to the correlation chip(s). This requires that the processor executes at a fast enough rate to read, interpret, and adjust the outputs and inputs of the signal correlator channels in real time. The conventional ASIC-based system has become a compact solution incorporated into most navigation systems available today.

However, pressure for improvements to this architecture comes principally from two directions. The first is the advancement of field-programmable gate array (FPGA) technology, which allows for hardware that could greatly expand the capability and flexibility of the receiver's digital components. For example, an FPGA could be programmed with hundreds of correlator channels, which could, in turn, be reprogrammed to suit individual application needs. The second impetus for improvement is the ever increasing speed of software processing. The software GNSS receiver has demonstrated that many of the functions of digital hardware can be performed in software. This has opened up the possibility of even greater flexibility in system design, while eliminating several cumbersome and often power-hungry components. However, the speed of software is still much lower than that of a hardware circuit, so this route often faces the formidable challenge of maintaining hardware-comparable performance. It is often the case that FPGA and software-based receivers are used during the design and prototyping phases of development to determine the optimal configuration for a system. Subsequently, the design is converted to an ASIC to better optimize the speed, cost, and power performance of the final receiver. Sections 1.3.2 and 1.3.3 briefly describe the FPGA-based receiver and the software receiver.

1.3.2 FPGA-Based Receivers

The FPGA-based receiver has grown significantly in popularity over the last decade, due primarily to the high circuit density achievable on a relatively small programmable chip at a low cost. The FPGA is programmed in a custom design language known as VHDL. When this code is loaded onto the chip, the millions of generic circuit gates are configured according to the programmed logic to perform the specified functions. This process can be viewed roughly as writing a computer program that is not compiled in the traditional sense to be run on a processor but is "compiled" to configure generic hardware gates on a chip. The advantage is that the resulting chip runs with the speed of hardware but maintains the ability to be programmed like software.

The disadvantage of the FPGA is that the hardware can only be efficiently programmed for certain tasks; in fact, many functions remain difficult to implement at all. For example, performing GPS signal tracking in an FPGA is not difficult as it consists primarily of a long series of multiplication and accumulation (functions well-suited to generic hardware gates). Alternatively, complicated control logic, which is often present in navigation software control architectures, is not efficient to perform in an FPGA; therefore, a separate control processor is still needed. Another task that is well-suited to FPGAs is performing fast Fourier transforms (FFTs), often used during a GNSS receiver's initial signal search and acquisition operations.

These two main benefits—fast and abundant parallel correlators and FFTs—need to be considered next to the increase in power and space that often come with an FPGA. The power issues need to be considered carefully as FPGAs can draw more power than their ASIC cousins, but with the devices becoming smaller and more power-efficient this issue can often be mitigated adequately as well.

1.3.3 Software-Defined GNSS Receivers

The last type of GNSS receiver coming into wide use is the software-defined radio or software receiver. The primary innovation of the software receiver is the elimination of digital processing hardware, resulting in a great simplification of the receiver design, including decreased size, power consumption, cost, and flexibility.

Unfortunately, there is a downside: The tasks performed by the digital correlators still need to be completed; they have just been moved into the control processor, thus greatly increasing the processing load on the software. The software is now required to perform a frequently large array of multiplications and accumulations at millisecond intervals. Even with the great advances in computing power over the past decades this increased burden still remains a challenge. However, it has been demonstrated that a software receiver can run with up to 12 channels in parallel in real time [13, 14], and it is expected that as processors become even faster, the computational burden of the software receiver will become easier to handle.

One frequently overlooked benefits of software receivers is that they force software and system designers into the innermost details of signal processing (places where only analog and digital hardware engineers have previously ventured), where limited visibility into the calculation was floated to the surface, and where there was no possibility of freezing, rewinding, or replaying the internal processing functions. With software, designers are permitted to process the same data over and over, trying new techniques and developing new tricks that are all under their complete control. This makes working with software receivers one of the best ways to learn about GNSS receivers, down to the finest details.

Independent of the argument about whether ever increasing processor speeds will make software receivers ubiquitous, or if their performance will always trail behind FPGA- and ASIC-based receivers, is the fact that the software-defined receiver will open up the techniques of GNSS receiver development to a new generation of designers. All one needs to get started is a computer running a software receiver and some sampled data on disk. Incidentally, this is the subject of Chapter 5, which presents a software receiver and two sample data sets that will allow the reader to explore the details of GNSS acquisition, tracking, and positioning algorithms.

1.4 Road Map: How To Use This Book

This book was written with the assumption that most readers will jump around from chapter to chapter depending on their interests. It is unlikely that the users of this book will read it from cover to cover. This is based on our experience of

how practicing engineers, scientists, and students use a book describing some given technology. A case in point is to consider how most people use software manuals!

Generally, a question will arise in the context of an ongoing project, and then a series of textbooks will be scanned, and colleagues consulted, in search of an answer or, even better, an example. A person looking to learn more about how a car navigation system works may have little or no interest in surveying or ocean sensing applications. This observation has had a significant impact on how this book has been organized. We expect readers to jump ahead to the specific problem of interest, relying on their past training to cover the basics, and then jump back to cover the missing fundamentals when required. Imagine the student working on an automotive Electronic Stability Control (ESC) system for a class project. The student's initial questions may be the following:

- What kind of GNSS information is available in automotive navigation systems?
- What is the quality of the measurements provided?
- What are the limitations?

This book is intended to help this student with the navigation-related questions pertaining to his or her project, which may or may not be centered on satellite navigation. The student would initially read the case study provided on automotive applications provided in Chapter 7, which would provide the answers to many of his or her questions, such as those listed above. Following this, the student might need more information on the fundamentals of the GNSS measurements, concerning coverage and availability. For this he or she could consult Chapter 3 for a brief overview and run some simulated examples for his or her particular problem. Should he or she want more details on the nature of the signals and measurements, it would then be necessary to consult an existing text that covers these topics more explicitly [1]. In this way students or engineers have the option of working backwards to the degree that is suitable for their project.

It is important to note that this book only provides brief but useful introductions to complicated fields of development and research. For this reason, some chapters include detailed references for those interested in pursuing these applications in greater depth. The chapters are all outlined below, including a brief chapter summary and an explanation of the examples (when available) included on the accompanying DVD.

Chapter 1 Global Navigation Satellite Systems: Present and Future

This chapter's main purpose is to provide some general context for the rest of the book. It includes general descriptions on the status of GNSS, navigation systems, receiver technology, a detailed outline (i.e., this roadmap), and a summary of the hands-on data provided as part of several chapters. The roadmap, including the summary of the data provided on the DVD, is intended to act as a guide to the contents of each chapter and the materials and subjects covered in each.

Chapter 2 GNSS Signal Acquisition and Tracking

This chapter provides a brief overview of how the signals transmitted from GNSS satellites are turned into measurements. The topics covered include: basic and novel

acquisition and tracking techniques for traditional (PSK) and new (BOC) GNSS signals.

Examples on the DVD. The software receiver discussed in Chapter 5 can be used to demonstrate many of the basic methods of GNSS receiver acquisition and tracking.

Chapter 3 GNSS Navigation: Estimating Position, Velocity, and Time

This chapter explains how to calculate the position, velocity, and time of a receiver. This will include a presentation of the least-squares position and velocity solution and an introduction to the extended Kalman filter (expanded on in Chapter 6) navigation methods. Chapter 3 also assesses the issue of measurement geometry and details the error sources with several examples.

Examples on the DVD. A GNSS simulator has been provided on the DVD that will allow the user to simulate unique scenarios on a global scale, providing reasonably realistic raw measurements and navigation solution calculations (repeatable using MATLAB/Octave and Python scripts). This program includes a GNSS constellation and receiver simulator that outputs a range of measurements under user-defined environments, which can then be used in conjunction with postprocessing scripts to duplicate the results presented in the text. This simulation can also be configured to provide outputs in a format compatible with Google Earth and to use time-tagged position inputs from FlightGear aircraft simulations. Finally, the software receiver discussed in Chapter 5 can be used for additional experimentation with the navigation calculations presented, using real data.

Chapter 4 Differential GNSS: Accuracy and Integrity

Chapter 4 discusses generic DGNSS systems and provides details on the use of space-based augmentation systems, such as the WAAS used extensively in North America.

Examples on the DVD. While there are no specific examples associated with this chapter, the software receiver discussed in Chapter 5 is capable of making WAAS ionosphere corrections as described in Chapter 4, thereby demonstrating a simple DGNSS application using real data.

Chapter 5 A GPS Software Receiver

A fully functional software receiver has been designed to demonstrate many of the steps performed in a GNSS navigation receiver, as overviewed in Figure 1.4 and Chapters 2–4. This receiver includes: a FFT acquisition; code and carrier tracking; pseudorange measurements; navigation data decoding; and position, velocity, and time estimation functions. The software receiver is also capable of using external aiding data from the International GNSS Service (IGS) as well as applying WAAS differential ionosphere corrections. It has been written using freely available open-source tools and is released under a GPL open-source license that will allow people to reuse, distribute, and upgrade the receiver to suit specific applications.

Examples on the DVD. The open-source fastgps software receiver is provided in several versions. This includes a basic Makefile terminal version as well as simple graphical user interface versions for Linux, Mac, and Windows. Also included are two example data sets, IGS aiding files and an example WAAS ionosphere differential correction file.

Chapter 6 Integration of GNSS and INS: Part I

This chapter discusses the integration of GNSS with INSs. In addition, Chapter 6 describes the motivation for integrating GNSS with INS and the various architectures used for this integration. It also discusses the advantages and disadvantages of these architectures and presents, in some detail, the algorithms used for the integration of GNSS with low-cost inertial sensors. The focus is on low-cost inertial sensors, because these are the sensors that are most readily available to users and developers. This chapter focuses on the theory of GNSS and INS integration and the following chapter deals with specific, hands-on examples.

Examples on the DVD. MATLAB functions for performing various routine INS calculations are provided.

Chapter 7 Integration of GNSS and INS: Part II

This chapter uses the theory developed in Chapter 6 to show the reader how to integrate GNSS with INS.

Examples on the DVD. An algorithm for the loose integration of GNSS and INS is included. A simulator for generating IMU is given. Data sets from flight tests of actual vehicles are also included. These data sets will allow the reader to design a GNSS/INS filter and develop experience with tuning and optimizing the sensor integration algorithms.

Chapter 8 Integrated LADAR, INS, and GNSS Navigation

Continuing on the theme of integrating GNSS with other sensors for performance enhancement, Chapter 8 explores the use of LADAR along with INS and terrain maps for this purpose. Chapter 8 provides a historical overview of this integration scheme and then discusses the architecture used for integrating GNSS, LADAR, INS, and terrain maps. The chapter culminates in the presentation of test results from a flight test campaign performed to validate the functionality and performance of such systems.

Chapter 9 Combining GNSS with RF Systems

We live in a world awash with RF signals, and they all contain some information that can be exploited for navigation. Chapter 9 explores the topic of combining GNSS with other RF systems for the navigation and tracking of vehicles, personnel, and assets. Chapter 9 first discusses the basics of RF signal propagation and then considers the motivation for combining various RF signals with GNSS. The chapter then provides examples of currently available RF signals that can be exploited for this application and presents filtering and sensor fusion algorithms for combining RF signals with GNSS. The chapter then closes by describing the applications that use the integration scheme discussed in this chapter.

Examples on the DVD. A spreadsheet and MATLAB/Octave scripts are provided that demonstrate basic RF antenna computations as well as Doppler and K-Nearest Neighbor estimation techniques.

Chapter 10 Aviation Applications

Aviation users represent a small fraction of the overall market for GNSS devices, but their demanding applications continue to advance the state of the art of satellite navigation technology. Chapter 10 focuses on augmentation systems as the cornerstone for aviation applications of GNSS. The first half of this chapter discusses the basic classes of augmentation systems and their applications. The second half describes quantitative techniques used to analyze augmentation system performance.

Chapter 11 Integrated GNSS and Loran Systems

Chapter 11 discusses the integration of GNSS with the low-frequency radio navigation system known as LORAN. It is well-known that GNSS are susceptible to interference and jamming. Thus, as more and more users start relying on GNSS for various applications, ensuring that it is robust against interference and jamming has become very important. LORAN has emerged as one potential candidate system that can be integrated with a GNSS to enhance its robustness. Chapter 11 describes LORAN signals and receivers to facilitate this integration (especially for safety-of-life applications) and discusses various integration architectures for these two systems and the trade-offs associated with these integration schemes. Finally, the issue of integrity is considered, which must be carefully considered when these two systems are used in applications with safety-of-life implications.

Chapter 12 Indoor and Weak Signal Navigation

Many modern satellite navigation applications are hampered by the degradation of signals in obstructed environments, such as in urban canyons, indoors, or within dense forests. There has been a significant amount of research into ways of overcoming the problems caused by this signal degradation. Chapter 12 gives an overview of the current state of this research and how it can be applied in practice. Included in this discussion are aiding and assistance methods, dealing with acquisition and tracking of very weak signals, as well as ways to perform measurement integrity monitoring.

Examples on the DVD. Two examples are included on the DVD using simulated GNSS measurements and MATLAB/Octave scripts. The first demonstrates the problems involved in acquiring signals at a very low power level. The second shows how to use statistical methods to detect erroneous measurements and to eliminate them from a modified least squares position solution.

Chapter 13 Space Applications

GNSS receivers have become a standard sensor on nearly all Earth orbiting satellites, and often they are used for more than just navigation. Chapter 13 provides an overview of the issues involved in adapting a GNSS receiver to operate in space and reviews the most common space-based applications. It covers several aspects

of space-based GNSS receiver operations and discusses the important issues that must be dealt with when operating a GNSS receiver on a spacecraft. Also discussed are diverse topics such as navigating with receivers in orbits above the GNSS constellation and determining a satellite's attitude with GNSS signals.

Examples on the DVD. A MATLAB/Octave script is included to estimate the position of the GRACE satellite using raw measurements over a single orbit. The measurements from the Blackjack dual-frequency GPS receiver onboard both GRACE satellites are freely available on the Internet, and an example day-long data set has been included on the DVD as a demonstration. The data provided has been augmented to include the transmitting satellite positions and clock biases accomplished using IGS ephemeris files, GRACE online processing tools, the orbit interpolation code by NOAA, and a Python script to pull everything together, all included on the DVD. The estimated positions have been compared with the precise positions provided as part of the GRACE data download.

Chapter 14 Geodesy and Surveying

Chapter 14 provides a succinct overview of the GNSS methods and applications used in the fields of geodesy and surveying. It summarizes the state of the art methods used in surveying and geodesy applications and the range of measurements and processing techniques currently in wide use. As described in the chapter, the difference between these two fields is primarily a question of required accuracy, and this leads to very different requirements and processing techniques. In fact, it is possible to process data using a worldwide GNSS reference infrastructure with a number of Internet-based postprocessing networks. This chapter details the critical aspects of this vast field, giving readers a thorough introduction and a solid foundation for starting their own surveying or geodesy projects.

Examples on the DVD. To give the reader a hands-on experience in using the Internet-based processing tools discussed in this chapter, RINEX data files with raw observation and navigation measurements taken from a Topcon GPS receiver have been included on the DVD. These files can be uploaded to a couple of the processing sites mentioned to generate precise position estimates, which are returned in the form of e-mail reports to the user.

Chapter 15 Atmospheric Sensing Using GNSS Occultations

Using GNSS signals to sense the Earth's atmosphere has become a very useful tool to environmental scientists in trying to better understand the global climate. Chapter 15 discusses the basic techniques used to perform GNSS radio occultation measurements using dual-frequency GNSS measurements from Low Earth Orbit (LEO) spacecraft. In addition, Chapter 15 reviews the basic principals underlying this unique application and describes how the processing is carried out as a function of the environmental parameter that one is trying to sense and with respect to the geometry over the descent and rise of GNSS signals being tracked through the Earth's atmosphere.

Examples on the DVD. MATLAB/Octave scripts are included on the DVD that will process numerous data sets from the GRACE satellites and make estimates of the temperature and pressure profiles over the duration of the signal tracking through the Earth's atmosphere.

Chapter 16 Remote Sensing Using Bistatic GNSS Reflections

In addition to sensing the atmosphere, as described in Chapter 15, it is also possible to use GNSS signals as remote sensing radars, in both altimetry or scatterometry configurations. Chapter 16 overviews the novel area of remote-sensing the environment using bistatically reflected GNSS signals. Basic topics discussed include: reflection geometry, signal processing, ocean scattering, and surface modeling theory and potential applications. In addition, this chapter presents the applications of ocean altimetry, ocean scatterometry (wind and wave sensing), land sensing, and ice sensing, using the latest results from aircraft and spacecraft experiments.

Examples on the DVD. Included are MATLAB/Octave scripts that are capable of calculating specular reflection points on the Earth's surface as well as complete system models for generating multidimensional reflected signal maps. Additionally, a C based software receiver is included together with three space-based data sets (one from the ocean, one over land, and a third over sea ice) to allow readers to pursue their own ideas using real bistatic GNSS reflections data.

Chapter 17 New Navigation Signals and Future Systems in Evolution

The United States' GPS satellites are no longer the only navigation satellites orbiting the Earth, nor are the modern GPS satellites the same as their trail-blazing predecessors. Soon the skies will be filled with numerous new signals from several independently operated constellations. Chapter 17 reviews many of the new systems and signals that have already or will soon appear and discusses the underlying theory and technical details of how the many new features have been designed. Of interest to many GNSS users is how all the new systems will interact—or possibly, interfere—with one another. With this in mind, this chapter includes a significant practical discussion that allows readers to assess the potential levels of interference between systems, and thus to explore mitigation methods.

1.5 Further Reading

For more detailed information on the basic principals and applications of GNSS navigation we recommend the following text books:

- *GPS: Signals, Measurements and Performance* (Second Edition), by Pratap Misra and Per Enge (www.gpstextbook.com).
- *Understanding GPS: Principals and Applications* (Second Edition), edited by Elliot Kaplan and Christopher Hegarty.
- *GPS: Theory and Applications*, (Two Volumes), edited by Bradford Parkinson, James Spilker, Per Enge, and Penina Axelrad.

References

- [1] Harigae, M., et al, “Abreast of the Waves: Open Sea Sensor to Measure Wave Height and Direction,” *GPS World*, May 2005.
- [2] Larson, K., et al, “Crustal Deformation Measurements in Guerrero, Mexico,” *Journal of Geophysical Research*, Vol. 109, B04409, 2004.
- [3] Wickert, J., et al, “Atmosphere Sounding by GPS Radio Occultation: First Results from CHAMP,” *Geophysical Research Letters*, Vol. 28, No. 17, p. 3263–3266, 2001.
- [4] Misra, P., and P. Enge, *GPS: Signals, Measurements and Performance*, Lincoln, MA: Ganga-Jamuna Press, 2001. www.gpstextbook.com.
- [5] British Broadcasting Corporation (BBC), “First Galileo Signals Received.” BBC Web site, 13 January 2006. (<http://news.bbc.co.uk/2/hi/science/nature/4610452.stm>)
- [6] Cao C., Guifei J., and M. Lou, “COMPASS Satellite Navigation System Development,” Stanford Center for Position Navigation and Time Symposium, 2008, <http://scpnt.stanford.edu/pnt/index.html>.
- [7] U.S. Government Executive Branch, Vice Presidential Initiative, January 25, 1999.
- [8] Kaplan, E. D., and C. J. Hegarty (editors), *Understanding GPS: Principles and Applications (Second Edition)*, Norwood, MA: Artech House, 2006.
- [9] Navstar GPS, Interface Specification, IS-GPS-200, Revision D, 7 December 2004, Navstar GPS Space Segment/Navigation User Interfaces.
- [10] Galileo Open Service, Signal in Space Interface Control Document (OS-SIS-ICD), Draft 0. European Space Agency, Galileo Joint Undertaking, 2006, available at <http://www.galileoic.org>.
- [11] Jaewoo, J., *High Integrity Carrier Phase Navigation Using Multiple Civil GPS Signals*, Ph.D. Dissertation, Stanford University, August 2000. Available at http://waas.stanford.edu/pubs/phd_pubs.html
- [12] An example of commonly used GPS chipsets can be found at www.zarlink.com.
- [13] Ledvina, B., et al, “Bit-Wise Parallel Algorithms for Efficient Software Correlation Applied to a GPS Software Receiver,” *IEEE Transactions on Wireless Communications*, Vol. 3, Issue 5, September 2004, pp. 1469–1473.
- [14] Akos, D. M., et al, “Real-Time Software Radio Receiver,” *Proc. of ION-GPS National Technical Meeting*, 2001.

GNSS Signal Acquisition and Tracking

Paul Blunt

2.1 Introduction

This chapter details the process of acquiring and tracking various types of GNSS signals. Detailed comparisons are drawn between different techniques and algorithms for both phase shift keying (PSK) and binary offset carrier (BOC) receivers considering both performance measures and the relative hardware impact. The content of this chapter is largely drawn from the theoretical work and GNSS receiver designs given in [1], which is included on the DVD accompanying this book for further reference. This chapter begins by describing the techniques currently employed to locate and acquire PSK signals and then reviews the established tracking techniques for PSK signals. A summary of the more recent theory required for acquiring and tracking BOC signals is given with comparisons of the various techniques. This chapter outlines the problems facing receiver designers using conventional techniques for tracking BOC signals and details the current progress in addressing these problems. For those interested, Chapter 17 provides additional details on several other new GNSS signals.

2.2 GNSS Signal Background

A GNSS is a spread spectrum system, where the spread signal occupies a bandwidth much greater than the rate of the data being transmitted [2]. This redundancy of bandwidth serves to suppress the detrimental effects of interfering signals and reduces the peak transmitted signal power levels to effectively hide the signal in background noise. The spreading technique denoted Direct Sequence Spread Spectrum (DS-SS) refers to a technique where a carrier wave is modulated by a data signal overlaid with a high frequency pseudorandom noise (PRN) spreading signal.

Figure 2.1 shows a conceptual diagram of a spread spectrum system, with the modulation onto the carrier omitted for simplicity. The narrow bandwidth, B_d of the data signal $d(t)$ is spread by a PRN code spreading signal, $a(t)$ of significantly higher bandwidth, B_s . The transmitted signal then passes through a channel, which applies additive noise, $n(t)$ and interfering signals, $i(t)$. A synchronized replica spreading code signal multiplied onto the received signal will then result in recovery

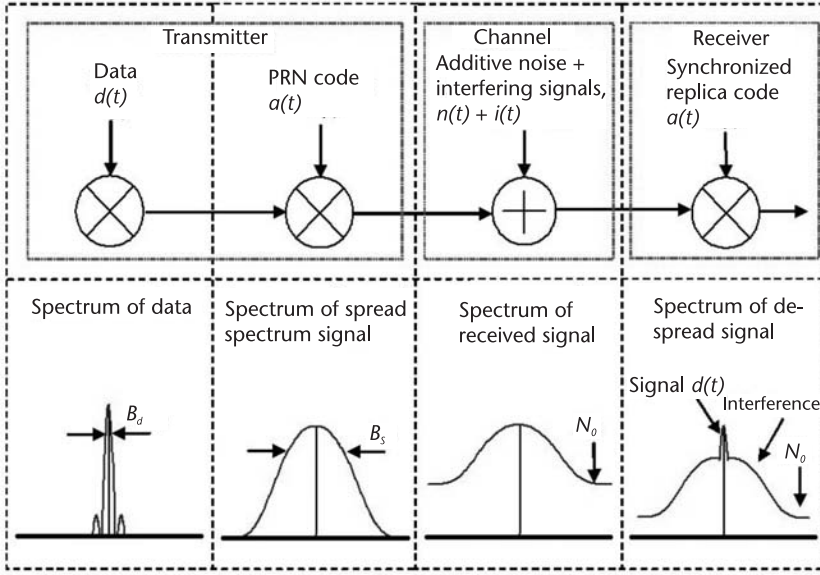


Figure 2.1 Simplified conceptual spread spectrum system.

of the desired signal with some error from thermal noise (spectral density N_0) and interference.

The innovation of GNSS, including the Global Positioning System (GPS), is that it uses the PRN code sequence as a ranging signal. In combination with its associated data signal this allows the path difference from transmitter to receiver to be recovered. The GPS satellite signals share the same carrier frequency and are separable in a receiver only because each respective transmission employs a unique PRN spreading code. Each effective bit of the PRN code sequence is called a *chip*. Despite the fact that GPS is a code modulation of a continuous wave (CW) carrier, the navigational or ranging signal can be viewed as a sequence of periodic pulses, with a periodicity equal to the code length and pulse width of one chip. The range to each differently located GPS satellite is measured by the timing of these pulses, and comparing the relative time delay of the pulses enables three-dimensional navigation.

Heritage GPS uses binary PSK (BPSK) modulation [3], where ideally the carrier phase changes instantaneously by 180° , depending on the data-modulated spreading sequence. A BPSK modulated signal can be written as

$$S_{PSK}(t) = \sqrt{2P} \times a(t) \times \cos(\omega t) \times d(t) \quad (2.1)$$

where $d(t), a(t) \in (-1, +1)$.

P is the signal power, $d(t)$ is the biphase data signal, $a(t)$ is the biphase PRN code spreading signal, and ω is the carrier frequency. The designation BPSK-R(f_c) has been adopted to define this type of modulation, where f_c is the chipping rate and is a multiple of 1.023 MHz.

Figure 2.2 shows the L-band frequency allocations for the current and future GPS and Galileo systems. Currently, GPS consists of between 24 and 32 satellites

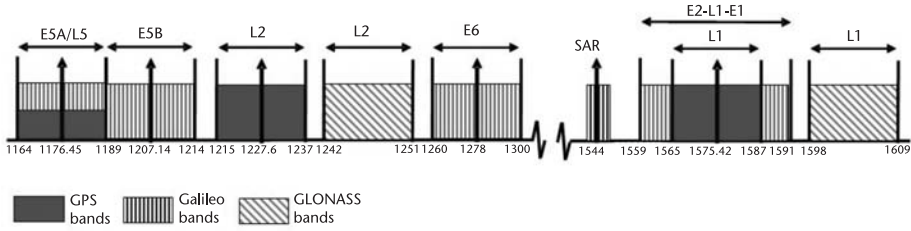


Figure 2.2 L-Band GNSS frequency allocations.

broadcasting three navigational signals in the L-band, one signal available for civil use transmitted on the L1 carrier frequency and two military signals transmitted on the L1 and L2 carriers. The future 30-satellite Galileo constellation will introduce six data-modulated and four dataless signals into the L-band spectrum occupying four frequency bands.

2.2.1 BOC Signal Modulation

A 12-month study by the GPS Joint Program Office (JPO) [4] resulted in the decision to use BOC modulation for the new military signals on L1 and L2. This modulation type will also be implemented for Galileo signals in various forms. BOC modulation is a rectangular subcarrier modulation (sine or cosine) of the PRN spreading code and is denoted $\text{BOC}(f_s, f_c)$, where f_s is the subcarrier frequency, and f_c is the PRN code chipping rate, and both are multiples of 1.023×10^6 . The subcarrier frequency is chosen such that it has an integer number of half periods, T_s (subchips) within a chip of the spreading sequence.

The effect of BOC modulation is that the signal is amplitude-modulated, with its power spectrum shifted away from the carrier frequency in upper and lower sidebands with a null at the center frequency (see Figure 2.3). The null is the primary benefit of BOC modulation, allowing frequency reuse alongside existing PSK signals. In addition, BOC modulation claims greater resistance to short-range multipath and a small advantage in code tracking accuracy. However, these advantages come at a price, namely the difficulties in acquiring and tracking signals with a subcarrier modulation.

A BOC signal is created through modulation of a PRN sequence, $a(t)$ by a square wave subcarrier, $s(t)$ and can be represented as follows.

$$S_{\text{BOC}}(t) = \sqrt{2P} \times b(t) \times \cos(\omega t) \times d(t) \quad (2.2)$$

where $b(t) = a(t) \times s(t)$. BOC signals in various forms are to be used by both current and future GNSS. The GPS M code signals are to be transmitted using a BOC (10, 5) modulation with a sine-phased subcarrier in the L1 and L2 bands. The Galileo signals for the E5a and E5b bands are combined onto a single-center frequency though the use of alternate BOC (AltBOC) modulation [5]. Recently, a joint GPS-Galileo working group [6] proposed a common multiplexed BOC (MBOC) structure for future L1 civil signals.

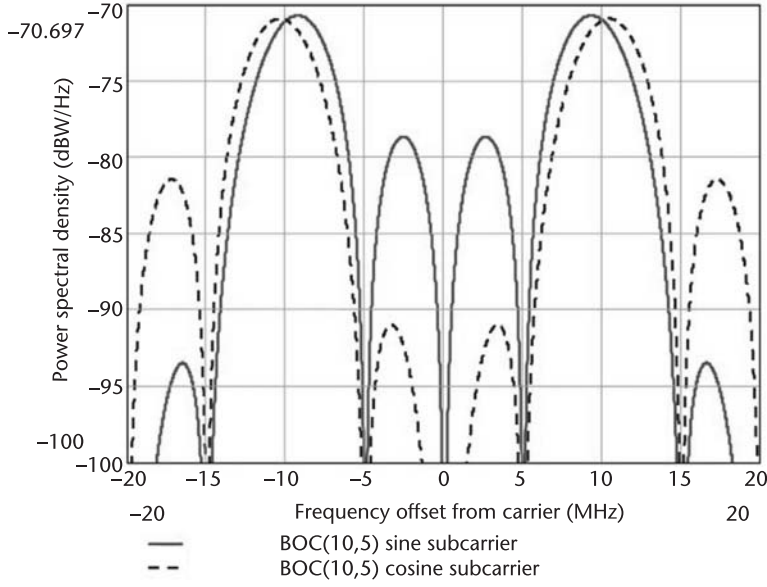


Figure 2.3 Power spectral density of a BOC(10, 5) signal with sine and cosine subcarriers.

The theory, techniques, and discussion presented in this chapter are intended to provide the basic receiver principles used in the reception of PSK and BOC signals. The reader is encouraged to consult the cited references for more information on special cases such as AltBOC and MBOC modulations. Additionally, Chapter 17 contains more information on these modulation techniques.

2.2.2 PRN Codes

The GPS C/A code is generated using two 10-bit linear feedback shift registers (LFSR) [3]. The resulting PRN codes are maximal length sequences with $N = 2^n - 1$ chips, where n is the length of the LFSR. The periodic autocorrelation function for a PRN sequence $a(t)$ of length n chips, with a chip period T_C can be written as follows.

$$\Lambda(\tau) = \frac{1}{NT_C} \int_0^{NT_C} a(t)a(t + \tau) dt \quad (2.3)$$

The autocorrelation envelope (Figure 2.4) can be approximated by a triangle function, the peak of which (amplitude A) corresponds to the perfect alignment (correlation) between the received code and the locally generated replica. Outside the correlation interval the ideal cross-correlation function is $-A/N$ due to the avoidance of the all-zero (stable) state in the generation of Gold codes.

In practice the sequences identified by Gold exhibit a three-valued cross-correlation function with values $\left\{ \frac{-1}{2^{n-1}}, \frac{-t(n)}{2^{n-1}}, \frac{t(n)-2}{2^{n-1}} \right\}$, where

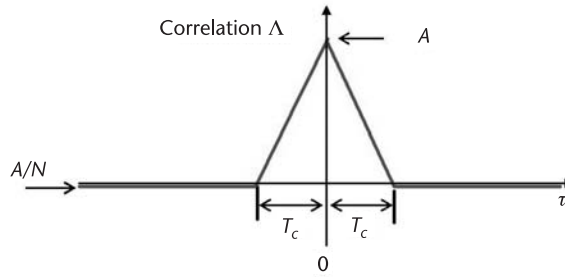


Figure 2.4 Ideal autocorrelation of a PRN sequence.

$$t(n) = \begin{cases} 2^{(n+1)/2} + 1 & (\text{odd } n) \\ 2^{(n+2)/2} + 1 & (\text{even } n) \end{cases} \quad (2.4)$$

Therefore, Gold sequences employed for GPS C/A code have cross-correlation values of $\left\{ \frac{-1}{1023}, \frac{-65}{1023}, \frac{63}{1023} \right\}$. The autocorrelation of the GPS C/A code (PRN1) is shown in Figure 2.5.

The modernized GPS signals use PRN codes generated by LFSRs as does Galileo along with precomputed memory codes. Conceived initially for the new GPS L5 signal, “tiered” PRN codes are to be broadcast by both GPS modernized and Galileo signals. A tiered system overlays the fast changing PRN sequence, which is called the primary code, with a secondary code whose bit period is equal to one primary code epoch. The secondary code length is then chosen such that the secondary code period is equal to one symbol period of the navigational data stream. The aim is to rapidly eliminate the ambiguity that initially exists in locating the PRN code epoch relative to the longer length data transitions.

A code or channel with no data modulation is known as a *pilot*; this is an important design feature and is incorporated many times into the modernized GPS and Galileo signal structures. A pilot code or channel allows GNSS receivers to use true phase-locked loop (PLL) discriminators, providing more accurate tracking

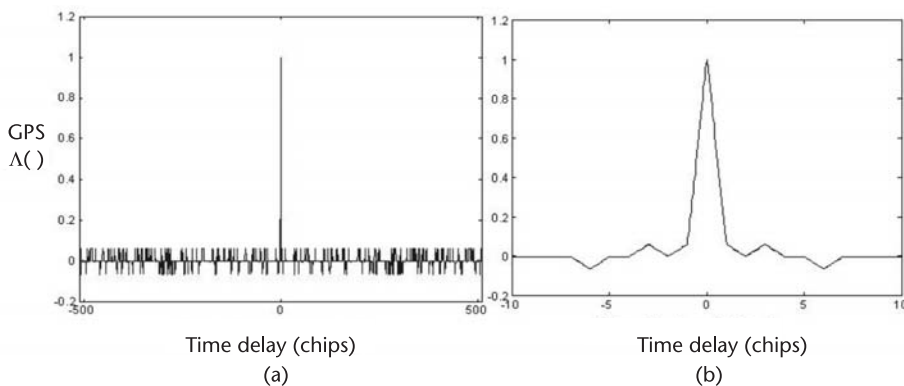


Figure 2.5 (a, b) Normalized autocorrelation of GPS C/A code.

than the Costas loop discriminators, which are commonly used in existing GNSS receivers. As discussed in Chapter 12, the pilot tones also allow longer integration periods to be easily implemented in the receiver for weak-signal acquisition and tracking. MATLAB/Octave and Python scripts, which generate several PRN codes, are included on the DVD (/Misc/GNSS_Code_Generators_ICDs/).

2.3 Searching for PSK Signals

To understand how GNSS signals are turned into navigation solutions it is necessary to gain a familiarity with the basic operation of GNSS receivers. This section describes the essential processes of a receiver's acquisition process, and Section 2.4 shows how the signal tracking is performed in a conventional GNSS receiver.

Figure 2.6 shows a generic block diagram for a GNSS receiver. Although there are a number of changes required to receive new-generation GNSS signals, the generic functions of the receiver are still valid.

GNSS signals are normally received through the use of a right-hand circularly polarized (RHCP) antenna and amplified using a low-noise amplifier (LNA), which essentially determines the receiver's noise figure. The RF signals are down-converted, typically in a number of stages, to an intermediate frequency (IF), sufficiently high in frequency to support the signal bandwidth.

The signal is then digitized by an analog-to-digital converter (ADC), with automatic gain control (AGC), and the digital IF is then passed to the receiver's correlator channels. Here, the carrier signal and code sequence are removed from the signal by correlating the received signal with locally generated replicas. The processor then extracts the raw navigational data by monitoring the changes in phase angle. This data can be used in combination with phase information derived from the carrier and code tracking loops to form pseudorange and Doppler estimates, ultimately resulting in position, velocity, and time information for the user. Chapter 5 explains this process in more detail, including a demonstration using a software receiver and sampled IF data.

The RF signals may also be directly mixed down to baseband or near-baseband frequencies prior to digitization. This technique can simplify RF front-end design

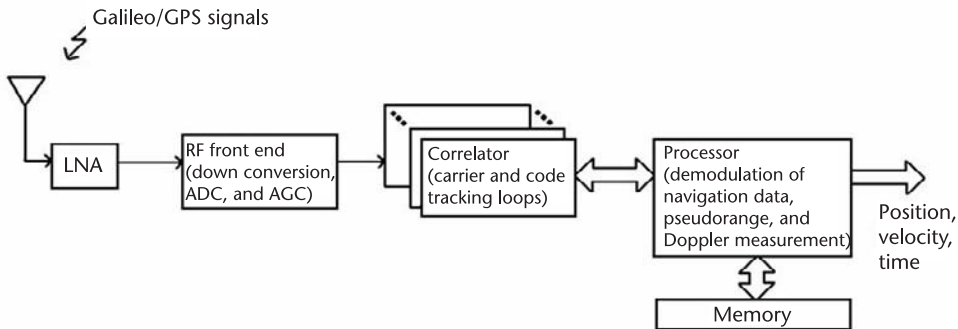


Figure 2.6 Simplified block diagram of a basic GPS receiver channel.

but requires careful design and treatment in the receiver; see [7] for an in-depth discussion.

Before GNSS signals can be used to solve for the receiver's position the signals must be found. The signals arriving at the receiver have an associated delay due to the distance between transmitter and receiver but have also been shifted in frequency. Therefore, the receiver must search in both frequency and delay (time) domains.

Assuming signal conditioning and down-conversion to a suitable IF we can model the received PSK signal as follows.

$$u_{PSK}(t) = A \times \cos(\omega_0 t + \phi) \times a(t - \tau) \times d(t) \quad (2.5)$$

A is the amplitude of the signal, ω_0 is the center frequency of the IF signal, $a(t)$ is the PRN code sequence, and $d(t)$ is the navigational data value having possible values of +1 and -1. τ is the time delay of the code, and ϕ is a general phase shift implicitly allowing for time variation in phase as follows.

$$\phi = -\omega_0 \tau + \phi_0 \quad (2.6)$$

ϕ_0 is an unrelated phase shift from the uncharacterized path from transmitter to receiver. Allowing a time-varying delay, $d\tau/dt$ accounts for the Doppler shift $d\phi/dt$ on the incoming signal.

Additive noise and interfering signals are not present in this representation to simplify the following illustrations. The aim of a PSK receiver is to estimate the delay, τ , and to demodulate the incoming navigational data. This is accomplished using estimating correlators (see Figure 2.7) and feedback (a digital PLL). The receiver produces replica in-phase and quadrature carrier signals with trial phase $\hat{\phi}$, which are mixed with the incoming signal. Subsequently, the signal is mixed with replica in-phase and orthogonal PRN codes with trial time delay $\hat{\tau}$. The

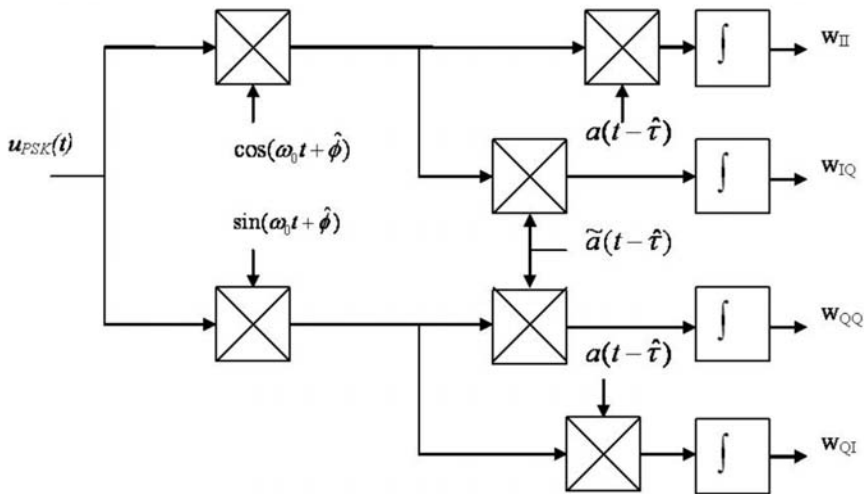


Figure 2.7 PSK correlator structure.

orthogonal code sequence is defined as the difference between *early* and *late* time shifts of the code sequence, written as follows.

$$\tilde{a}(t - \hat{\tau}) = a\left(t - \hat{\tau} + \frac{T_{DC}}{2}\right) - a\left(t - \hat{\tau} - \frac{T_{DC}}{2}\right) \quad (2.7)$$

T_{DC} is the total separation between the *early* and *late* replica waveforms, bounded by $T_{DC} \leq T_C$. Multiplication of the incoming signal by carrier and code replicas results in four correlation results denoted w_{II} , w_{QI} , w_{IQ} , and w_{QQ} shown in Figure 2.7. The first subscript denotes mixing with an in-phase (*I*) or quadrature (*Q*) carrier replica; the second subscript denotes mixing with in-phase or quadrature (orthogonal) code replica. The w_{II} and w_{QI} correlations are used for the receiver search process and carrier tracking. The additional w_{IQ} and w_{QQ} correlations are used along with the w_{II} and w_{QI} correlations to track the PRN code within the received signal (see Section 2.4).

The resulting correlations can be written as follows. For more details on this derivation see [1].

$$\begin{aligned} w_{II} &\approx A \times \cos(\phi - \hat{\phi}) \times \Lambda(\hat{\tau} - \tau) \times d \\ w_{IQ} &\approx A \times \cos(\phi - \hat{\phi}) \times \sqrt{\Lambda}(\hat{\tau} - \tau) \times d \\ w_{QI} &\approx A \times \sin(\phi - \hat{\phi}) \times \Lambda(\hat{\tau} - \tau) \times d \\ w_{QQ} &\approx A \times \sin(\phi - \hat{\phi}) \times \sqrt{\Lambda}(\hat{\tau} - \tau) \times d \end{aligned} \quad (2.8)$$

The “ $\sqrt{\Lambda}$ ” represents the code tracking discriminator error signal, which is described fully in Section 2.4. The symbol “ Λ ” represents the ideal PSK correlation function, which can be written as

$$\begin{aligned} \Lambda(\hat{\tau} - \tau) &= \begin{cases} 1 - \frac{|\hat{\tau} - \tau|}{T_C} & \text{for } -T_C \leq (\hat{\tau} - \tau) \leq T_C \\ 0 & \text{otherwise} \end{cases} \\ &\approx \frac{1}{T} \int_0^T a(t - \tau) a(t - \hat{\tau}) dt \end{aligned} \quad (2.9)$$

The $\Lambda(\)$ function approximates the (almost) triangular correlation function of PSK PRN codes, ignoring cross-correlation errors. This function peaks as $\hat{\tau} \rightarrow \tau$, with total width we define as $2 \times T_C$. This can be seen by inspecting the result w_{II} derived from Mathcad simulation, which is shown in Figure 2.8.

For PSK signals the result of w_{II} is a single correlation peak; however the correlation still has a phase dependency and may be inverted with the sign of the navigational data. At startup the GNSS receiver must search for each signal with no knowledge of the phase of the received carrier. Phase independency for the PSK search correlation is achieved by adding the square of the w_{II} correlation combined

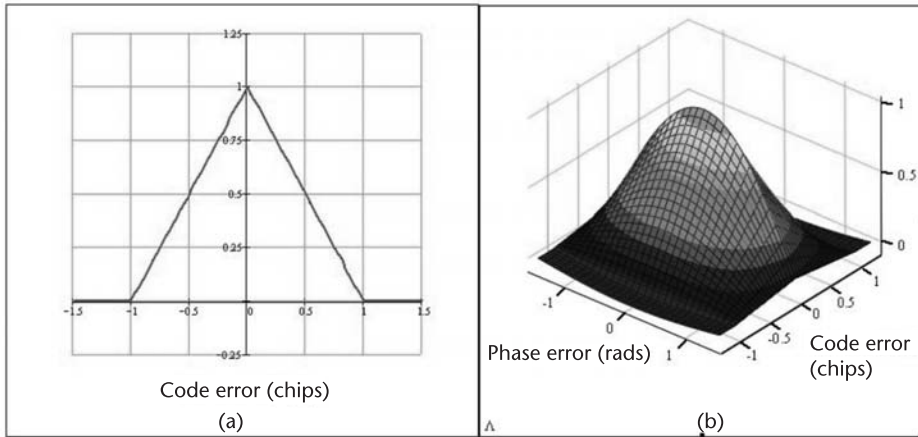


Figure 2.8 (a, b) Correlation profile of w_{II} against code error (chips) and phase error (radians).

with the square of the quadrature w_{QI} correlation. This forms an unambiguous peak for the search process, which is independent of phase error and the sign of the navigational data ($d^2 = 1$).

$$w_{SPSK} = w_{II}^2 + w_{QI}^2 \approx 2A^2\Lambda^2(\hat{\tau} - \tau) \quad (2.10)$$

The search is a two-dimensional process, locating the signal in time (code delay) and frequency. Traditionally, a serial search technique is used, holding a steady frequency while shifting through all possible code offsets, before moving to the next possible frequency (Figure 2.9). The frequency range the receiver is required

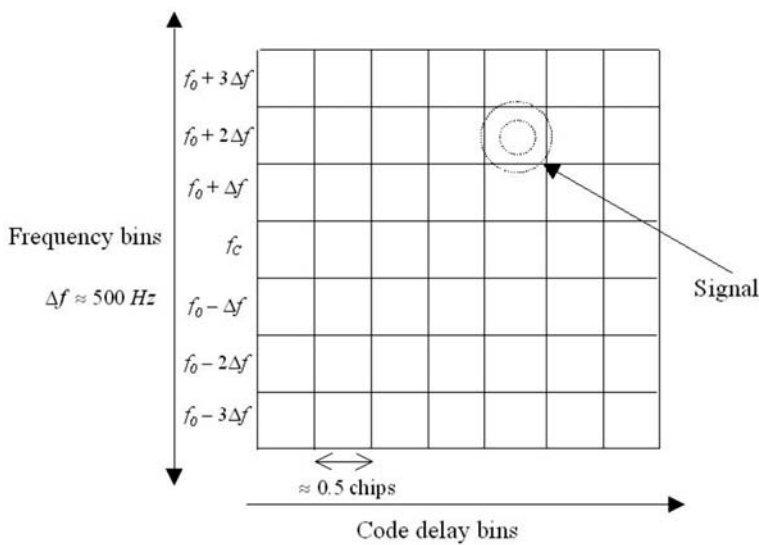


Figure 2.9 Serial GNSS signal search.

to search across is strongly dependent on the dynamic environment the receiver is to operate in and the quality of the receiver clock. For a typical terrestrial receiver the frequency search range can be up to ± 10 kHz; however for some applications such as space GNSS receivers the range can expand up to ± 50 kHz [8].

Frequency is typically searched in bins of 500 Hz (for $T = 1$ ms). The normalized correlation gain, G_F with frequency error, Δf is given by [9]

$$G_F = \frac{\sin(\pi \times \Delta f \times T)}{(\pi \times \Delta f \times T)} \quad (2.11)$$

where T is the integration period. Figure 2.10 shows the correlation gain for an integration time of 1 msec. The minimum correlation gain for frequency bins of 500 Hz is -0.9 dB. Extending the integration period allows detection of weaker signals. However, if the frequency bins are narrowed, it will result in longer search times and, potentially, reductions in correlation gain due to changes in the Doppler during the correlation time.

The code is typically searched in $1/2$ chip bins. Within the correlation interval ($\hat{\tau} - \tau \leq \pm T_C$) the normalized correlation power G_τ , with code delay error, $\hat{\tau} - \tau$ is given by

$$G_\tau = \Lambda^2(\hat{\tau} - \tau) \quad (2.12)$$

For $1/2$ -chip code bins the minimum correlation gain is -2.5 dB. Narrower code and carrier bin width will result in more reliable signal detection. However, the search time will increase because there are now more bins to search through.

The serial search technique is inherently slow and may take a number of minutes to detect signal presence reliably. Modern receivers often use search techniques

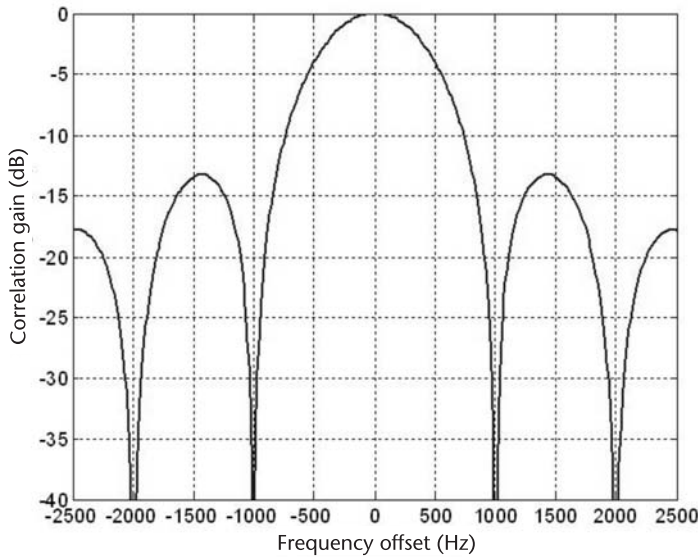


Figure 2.10 Correlation power loss with frequency offset.

based on the Fast Fourier Transform (FFT), which can detect the signal presence within a matter of milliseconds [10]. The basic principle of FFT signal detection is shown in Figure 2.11, first presented in [11]. The received signal is mixed into real (in-phase) and imaginary (quadrature) components and its Fourier transform computed. The result is then the conjugate multiplied by the Fourier transform of the code sequence. Multiplication of signals in the frequency domain is equivalent to correlating signals in the time domain. Therefore, the correlation across all code offsets can be computed by taking the inverse Fourier transform.

Figure 2.12 shows the FFT detection of a GPS C/A code signal. The result of a single FFT detection can deliver the equivalent to an entire serial search in a

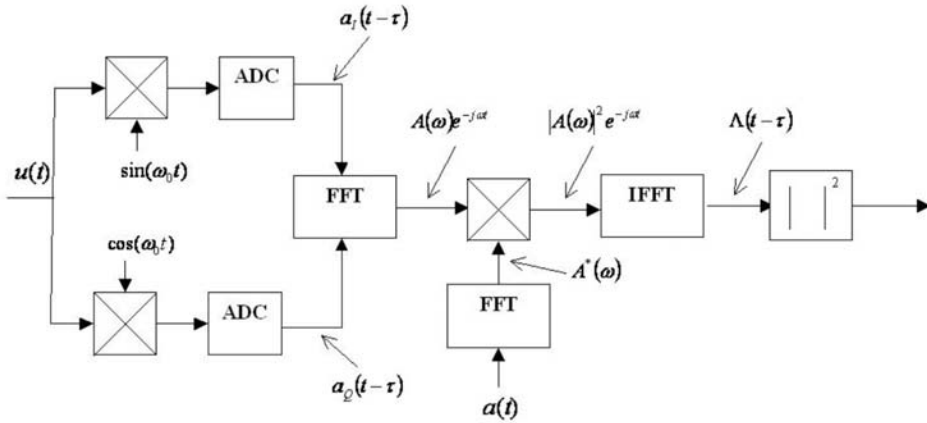


Figure 2.11 FFT signal detection principle.

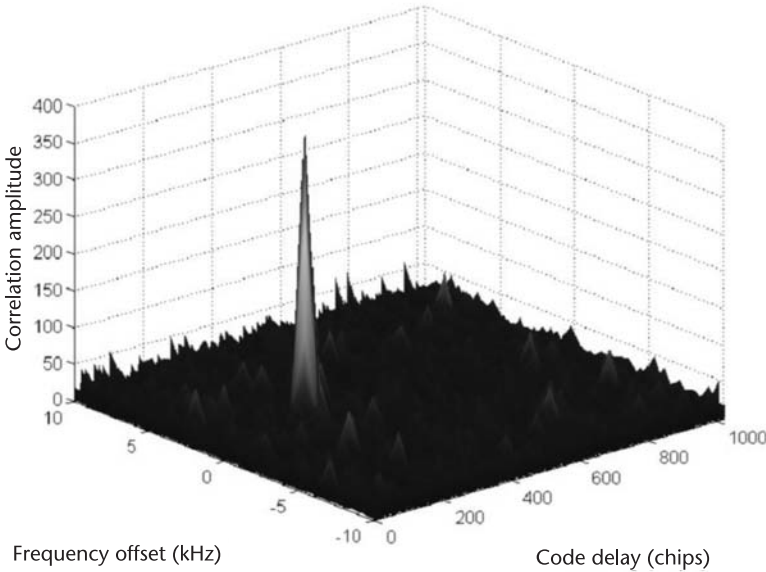


Figure 2.12 FFT of GPS C/A code signal.

matter of milliseconds. It is shown in [10] that FFT detection can be up to 2,000 times faster than the equivalent serial search technique. However, performing FFT detection in the receiver hardware imposes large processing overhead and often requires a dedicated FFT unit in the receiver design.

Whether using a serial search technique or FFT detection the correlation profile of the PSK signal is a single peak with correlation gains with respect to frequency and code error as given in (2.11) and (2.12), respectively. A detection threshold must be set within the receiver software above which signal presence is declared. This may be based on a single search result, *single dwell* or multiple search results, *multiple dwells*. Taking multiple dwells clearly increases the reliability of detection but also lengthens the signal search time. Raising the detection threshold reduces the risk of false alarm but also reduces the probability of detecting the signal. The choice of detection threshold is strongly dependent on the receiver's operating environment and application.

The software receiver presented in Chapter 5 uses an FFT-based signal acquisition strategy, where a 4-ms dwell time will result in four distinct peaks for a successfully detected 1-ms GPS C/A code PRN sequence. The reader is encouraged to experiment with this technique, including configuring the software receiver to run in debug mode, where the raw FFT acquisition outputs can be inspected individually (See Chapter 5 for details).

2.4 Tracking PSK Signals

To achieve a reasonable estimate of the time delay of the received signal, a PSK GNSS receiver must implement a carrier loop to maintain lock on the frequency and/or phase of the incoming carrier wave. Simultaneously, a code loop must also maintain lock on the PRN code sequence present in the received signal. This can only be achieved by adjusting the carrier phase estimate, $\hat{\phi}$ and the code phase estimate, $\hat{\tau}$ to track the incoming ϕ and τ respectively.

For the carrier loop it is possible to use a frequency-locked loop (FLL) or a PLL, from which we define an *incoherent* or *coherent* system, respectively. A FLL delivers more robust tracking in high-dynamic environments and in very weak signal conditions [12]. However, in order to enable precise pseudorange rate, integrated Doppler measurements and carrier phase positioning a GNSS receiver must track the phase of the incoming carrier. Hence, a coherent PLL system is required.

2.4.1 Phase-Locked Loop (PLL)

The carrier phase error is commonly determined using a Costas decision-directed discriminator [12] as follows.

$$e_{\phi} = w_{QI} \times \text{sgn}(w_{II}) \approx A \times \sin(\phi - \hat{\phi}) \times \text{sqc}(\phi - \hat{\phi}) \times \Lambda(\hat{\tau} - \tau) \quad (2.13)$$

where $\text{sqc}(\)$ is a square-wave cosine function, shown in the Figure 2.13. Incorporating the sign correction by hard-limiting the w_{II} correlation in the discriminator allows the PLL to function with no knowledge of the navigational data state.

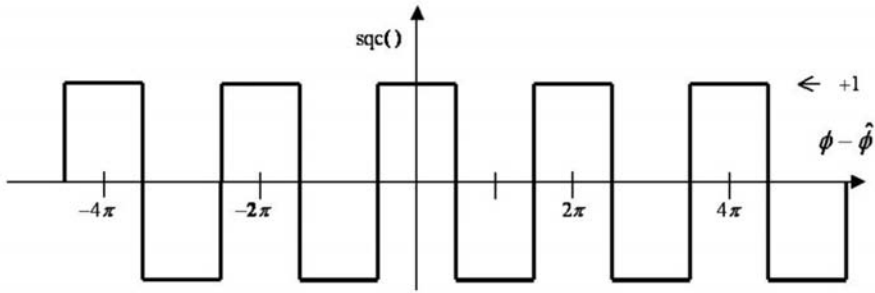


Figure 2.13 Square-wave cosine function.

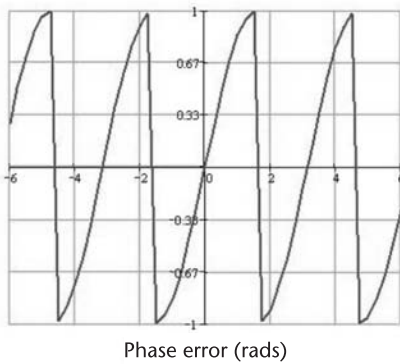
It can be seen that e_ϕ tends to zero as $\sin(\phi - \hat{\phi})$ tends to zero, which occurs at integer n multiples of π radians as $\hat{\phi} \rightarrow \phi + n\pi$. This 180° ambiguity is caused by a sign correction in the discriminator and can be resolved by inspection of the incoming navigational data. Figure 2.14 shows the discriminator characteristic.

The decision-directed discriminator is dependent on amplitude A . A number of PLL discriminators can be used, each with different dependencies and computational loads. Table 2.1 shows the most commonly used PLL discriminators and their error signals and dependencies. A receiver tracking a dataless *pilot* signal may use a pure PLL discriminator; see [12] for details.

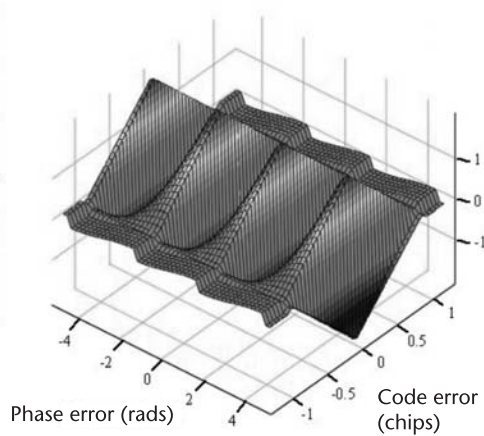
The PLL is generally updated using a second-order loop as follows.

$$\begin{aligned} f_\phi &\leftarrow f_\phi + e_\phi \\ \hat{\phi} &\leftarrow \hat{\phi} + k_1 \times f_\phi + k_2 \times e_\phi \end{aligned} \quad (2.14)$$

where f_ϕ is the integrated phase error, and k_1 and k_2 are loop gains that can be adjusted by the designer. More detailed treatments of how to set the gains for typical loops can be found in [1, 13].



(a)



(b)

Figure 2.14 (a, b) Costas PLL discriminator characteristic.

Table 2.1 Costas PLL Discriminators

<i>Discriminator</i>	<i>Error Signal</i>	<i>Dependency</i>
Decision-corrected	$e_\phi = w_{QI} \times \text{sgn}(w_{II}) \approx A \times \sin(\phi - \hat{\phi}) \times \text{sqc}(\phi - \hat{\phi}) \times \Lambda(\hat{\tau} - \tau)$	Slope proportional to A
Dot product	$e_\phi = w_{QI} \times w_{II} \approx \frac{A^2}{2} \times \sin[2 \times (\phi - \hat{\phi})] \times \Lambda^2(\hat{\tau} - \tau)$	Slope proportional to A^2
Normalized	$e_\phi = \frac{w_{QI}}{w_{II}} \approx \tan(\phi - \hat{\phi})$	Slope independent of amplitude
Two-quadrant arc tangent	$e_\phi = \arctan\left(\frac{w_{QI}}{w_{II}}\right) \approx (\phi - \hat{\phi})$	Slope independent of amplitude

2.4.2 Frequency-Locked Loop (FLL)

An incoherent system holds frequency lock by maintaining a constant or near constant phase difference across the correlation interval. The rate of change of carrier phase can be determined by comparing w_{II} and w_{QI} with the correlation results from the next epoch w'_{II} and w'_{QI} . The frequency error can then be determined using the cross-product discriminator as follows.

$$\begin{aligned}
 e_\omega &= w'_{QI} \times w_{II} - w'_{II} \times w_{QI} \\
 &= A^2 \times \sin(\Delta\phi - \Delta\hat{\phi}) \times \Lambda^2(\hat{\tau} - \tau) \times d \times d'
 \end{aligned} \tag{2.15}$$

The phase difference and estimated phase difference between epochs is

$$\begin{aligned}
 \Delta\phi &= \phi' - \phi \\
 \Delta\hat{\phi} &= \hat{\phi}' - \hat{\phi}
 \end{aligned} \tag{2.16}$$

As e_ω tends to zero, $\Delta\phi - \Delta\hat{\phi}$ tends to zero. This discriminator only gives a true frequency error when no data transitions have occurred. If correlations are unsynchronized with the data FLL discriminators only perform well when short integration periods are used to restrict the number of data transitions affecting the error characteristic. Filtering of the discriminator output may also be required to reduce the impact of the data transitions. It is common to use a second-order loop to update the carrier phase estimate by the frequency error as follows.

$$\begin{aligned}
 e_\phi &\leftarrow e_\phi + e_\omega \\
 f_\phi &\leftarrow f_\phi + e_\phi \\
 \hat{\phi} &\leftarrow \hat{\phi} + k_1 \times f_\phi + k_2 \times e_\phi
 \end{aligned} \tag{2.17}$$

e_ϕ is the estimated phase error derived by integrating the frequency error e_ω , k_1 and k_2 are loop gains which can be adjusted by the designer. The cross-product discriminator has a dependence on A^2 ; it is also possible to use a number of other FLL discriminators, each with different dependencies and computational loads.

Table 2.2 shows the most commonly used FLL discriminators and their error signals and dependencies.

2.4.3 Delay-Locked Loop (DLL)

Common practice is for the PRN code sequence to be acquired and tracked using a delay-locked loop (DLL). This can be achieved when implementing either incoherent FLL or coherent PLL carrier tracking. However, different correlations and discriminators are required for each system. A common discriminator suitable for tracking with both coherent and incoherent systems is the dot-product discriminator, formed as follows.

$$e_{\tau} = w_{IQ} \times w_{II} + w_{QI} \times w_{QQ} = A^2 \times \sqrt{\Lambda}(\hat{\tau} - \tau) \times \Lambda(\hat{\tau} - \tau) \quad (2.18)$$

The ' $\sqrt{\Lambda}$ ' symbol represents the code-tracking error equivalent to subtracting separate early and late correlations, written as follows.

$$\sqrt{\Lambda}(\hat{\tau} - \tau) = \Lambda\left(\hat{\tau} - \tau - \frac{T_{DC}}{2}\right) - \Lambda\left(\hat{\tau} - \tau + \frac{T_{DC}}{2}\right) \quad (2.19)$$

Figure 2.15 shows the dot-product discriminator characteristic. It can be seen that this discriminator has no phase dependence and is therefore suitable for an incoherent system.

It is common to use a first-order loop to update the code phase estimate. This is only made possible by the use of the Doppler aiding from the carrier-tracking loop. The carrier loop is generally a second- or third-order loop that accurately tracks the dynamics of the system. Therefore, the system dynamics can be effectively removed by applying the Doppler estimated by the carrier loop to the DLL with appropriate scaling. The first-order code loop update equation with carrier Doppler aiding can be written as follows.

Table 2.2 FLL Discriminators

<i>Discriminator</i>	<i>Error Signal</i>	<i>Dependency</i>
Cross-product	$e_{\omega} = w'_{QI} \times w_{II} - w'_{II} \times w_{QI}$ $\approx A^2 \times \sin(\Delta\phi - \Delta\hat{\phi}) \times \Lambda^2(\hat{\tau} - \tau) \times d \times d'$	Slope proportional to A^2
Normalized cross-product	$e_{\omega} = \frac{w'_{QI} \times w_{II} - w'_{II} \times w_{QI}}{w'_{II} \times w_{II} + w'_{QI} \times w_{QI}}$ $\approx \tan(\Delta\phi - \Delta\hat{\phi})$	Slope independent of amplitude
4 quadrant arc tangent (cross-product)	$e_{\omega} = \arctan 2 \left(\frac{w'_{QI} \times w_{II} - w'_{II} \times w_{QI}}{w'_{II} \times w_{II} + w'_{QI} \times w_{QI}} \right)$ $\approx (\Delta\phi - \Delta\hat{\phi})$	Slope independent of amplitude

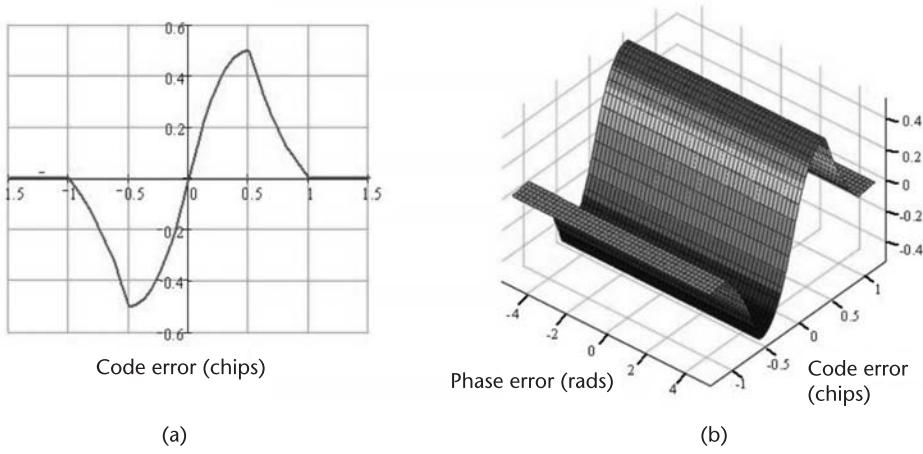


Figure 2.15 (a, b) Dot-product discriminator characteristic.

$$\hat{\tau} = \hat{\tau} + k_{\phi} \times f_{\phi} + k_{\tau} \times e_{\tau} \quad (2.20)$$

k_{ϕ} is a constant calculated to provide the necessary open loop correction of Doppler shift appropriately scaled down to the code rate and k_{τ} is a loop gain that can be adjusted by the designer. The dot-product discriminator is dependent on amplitude A^2 ; it is also possible to use a number of other DLL discriminators, each with different dependencies and computational loads. A number of discriminators used to remove the amplitude dependence require generation of signals individually multiplied by early and late replica code sequences as follows.

$$\begin{aligned} v_{IE}(t) &= v_I(t) \times a \left(t - \hat{\tau} \pm \frac{T_{DC}}{2} \right) \\ v_{IL}(t) &= v_I(t) \times a \left(t - \hat{\tau} \pm \frac{T_{DC}}{2} \right) \\ v_{QE}(t) &= v_Q(t) \times a \left(t - \hat{\tau} \pm \frac{T_{DC}}{2} \right) \\ v_{QL}(t) &= v_Q(t) \times a \left(t - \hat{\tau} \pm \frac{T_{DC}}{2} \right) \end{aligned} \quad (2.21)$$

The resulting early and late correlations are written as

$$\begin{aligned} w_{IE} &\approx A \times \cos(\phi - \hat{\phi}) \times \Lambda \left(\hat{\tau} \mp \frac{T_{DC}}{2} - \tau \right) \\ w_{IL} &\approx A \times \cos(\phi - \hat{\phi}) \times \Lambda \left(\hat{\tau} \mp \frac{T_{DC}}{2} - \tau \right) \\ w_{QE} &\approx A \times \sin(\phi - \hat{\phi}) \times \Lambda \left(\hat{\tau} \mp \frac{T_{DC}}{2} - \tau \right) \\ w_{QL} &\approx A \times \sin(\phi - \hat{\phi}) \times \Lambda \left(\hat{\tau} \mp \frac{T_{DC}}{2} - \tau \right) \end{aligned} \quad (2.22)$$

Table 2.3 shows the most commonly used DLL discriminators and their error signals and dependencies. The choice of discriminator must be made depending on the computational capability of the receiver and the signal environments in which it will operate. Removing the sensitivity of the receiver to signal amplitude provides the most robust solution but maximizes the microprocessor loading. Discriminators may also be normalized by the receiver's estimate of the carrier to

Table 2.3 DLL Discriminators

<i>Discriminator</i>	<i>Error Signal</i>	<i>Dependency</i>
<i>Coherent</i>		
Dot product	$e_\tau = w_{IQ} \times w_{II} \approx A^2 \times \cos^2(\phi - \hat{\phi}) \times \sqrt{\Lambda}(\hat{\tau} - \tau) \times \Lambda(\hat{\tau} - \tau)$	Slope proportional to A^2
Decision-directed	$e_\tau = w_{IQ} \times \text{sgn}(w_{II}) \approx A \times \cos(\phi - \hat{\phi}) \times \sqrt{\Lambda}(\hat{\tau} - \tau)$	Slope proportional to A
<i>Incoherent</i>		
Dot product	$e_\tau = w_{IQ} \times w_{II} + w_{QI} \times w_{QQ} \approx A^2 \times \sqrt{\Lambda}(\hat{\tau} - \tau) \times \Lambda(\hat{\tau} - \tau)$	Slope proportional to A^2
Decision-directed	$e_\tau = w_{IQ} \times \text{sgn}(w_{II}) + w_{QI} \times \text{sgn}(w_{QQ})$ $\approx A \times (\cos(\phi - \hat{\phi}) + \sin(\phi - \hat{\phi})) \times \sqrt{\Lambda}(\hat{\tau} - \tau)$	Slope proportional to A
Power	$e_\tau = w_{IE}^2 + w_{QE}^2 - w_{IL}^2 - w_{QL}^2 \approx A^2 \times \sqrt{\Lambda}(\hat{\tau} - \tau)$	Slope proportional to A^2
Normalized envelope	$e_\tau = \frac{\sqrt{w_{IE}^2 + w_{QE}^2} - \sqrt{w_{IL}^2 + w_{QL}^2}}{\sqrt{w_{IE}^2 + w_{QE}^2} + \sqrt{w_{IL}^2 + w_{QL}^2}}$ $\approx \frac{\sqrt{\Lambda}(\hat{\tau} - \tau)}{\left(\Lambda\left(\hat{\tau} - \frac{T_{DC}}{2} - \tau\right) + \Lambda\left(\hat{\tau} + \frac{T_{DC}}{2} - \tau\right)\right)}$	Slope independent of amplitude

noise density of the signal (derived from the prompt code correlations); see [13] for details. This approach avoids occasionally large error signals in weak signal conditions caused by division close to zero in the discriminator and removes any need for limiting the error signal.

Although listing discriminators may seem an exhaustive task, it is essential that receiver designers account for incoming signal amplitude variation in the loop designs. These tables provide designers with essential information for providing robust tracking loop design. For equivalent tables for BOC signal tracking please consult [1].

The performance of the tracking-loop discriminators described in Tables 2.1–2.3 can be demonstrated using the software receiver provided in Chapter 5. The default FLL and PLL configurations in the software receiver include dynamically switching the bandwidth of the FLL during frequency *pull in* before switching to a PLL for phase tracking and data demodulation. Readers are encouraged to explore the many options outlined above using the tools on the DVD, included as part of Chapter 5.

2.5 Searching for BOC Signals

A BOC-modulated signal is a square-wave subcarrier modulation of the coding sequence and can be written as

$$\mathbb{W}(\hat{\tau} - \tau) = \mathbb{W}\left(\hat{\tau} - \tau - \frac{T_{DC}}{2}\right) - \mathbb{W}\left(\hat{\tau} - \tau + \frac{T_{DC}}{2}\right) \quad (2.27)$$

Figure 2.17 shows the w_{II} correlation for a BOC $(2 \times f_C, f_C)$ -modulated signal. In order to locate the BOC signal the search process must remove the dependency on carrier phase error. This can be accomplished in the same manner as the conventional PSK search using the w_{II} and w_{OI} correlations as shown in (2.10). However, the multiple peaks and troughs of the BOC correlation function introduce nulls across the correlation interval, which reduce the probability of detection. For PSK 1/2-chip code search bins result in a minimum correlation gain of -2.5 dB from the peak gain. For BOC systems a minimum correlation gain of -2.5 dB is only achieved with search bins of half a subchip.

When implementing a standard serial search technique, BOC modulation increases the search time compared to PSK by approximately twice the ratio of subcarrier frequency to code rate [a factor of 4 for BOC $(2 \times f_C, f_C)$, or a factor of 12 for BOC $(6 \times f_C, f_C)$]. When FFT acquisition is employed, BOC modulation has an equivalent impact on the number of points required. Again, compared to PSK, the number of FFT points required to achieve equivalent correlation loss increases by twice the ratio of subcarrier frequency to code rate. Along with increasing code lengths and sampling rates this poses an unacceptable overhead on the receiver hardware. Therefore, for reliable fast acquisition of BOC signals it is necessary to form a single peak across the correlation interval for the search process.

Creating a single correlation peak for BOC search requires additional receiver hardware. To achieve this we consider two different approaches proposed in the literature, each with different receiver hardware requirements.

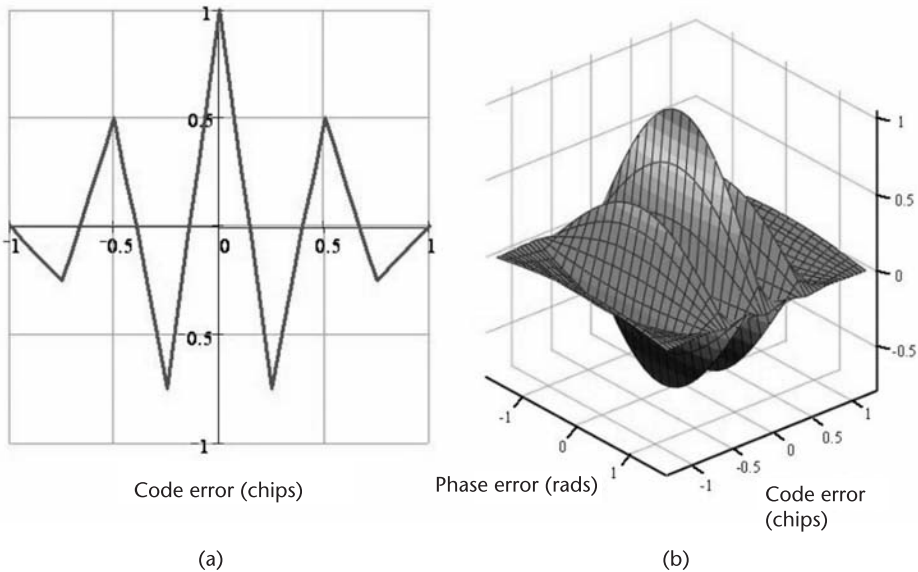


Figure 2.17 (a, b) w_{II} for BOC $(2 \times f_C, f_C)$ against code error (chips) and carrier-phase error (rads).

The first technique, single sideband (SSB) acquisition, was first proposed by Betz in [4]. This method treats each of the two BOC sidebands as separate PSK signals, requiring independent filtering of each sideband. In addition, each sideband must have a separate carrier demodulation stage, requiring an additional local oscillator to be implemented in the receiver's correlator architecture [1].

The result of an SSB search is an unambiguous PSK correlation peak, as shown in Figure 2.8. If only one sideband is tracked, a 3-dB or greater reduction in signal power is inevitable. The signal power loss can be compensated for by applying the SSB technique to each sidelobe and combining noncoherently [14]. This technique can easily be implemented for BOC signals that are well separated from the center frequency but that would require an extremely sharp filter roll-off (Nyquist filtering) for narrowly spaced signals, such as BOC (f_C, f_C) . Therefore, the SSB technique is only suitable for BOC signals whose subcarrier frequency is greater than the code rate.

The second approach to providing a single BOC search is to synthesize an unambiguous search function by using additional quadrature or orthogonal BOC correlations. This search technique first proposed by Ward in [15] has subsequently been coined the *subcarrier cancellation* (SCC) technique [16]. The BOC search correlation using the SCC technique can be written as follows.

$$w_{\text{SBOC}} = \sqrt{w_{\text{III}}^2 + w_{\text{QII}}^2} + \sqrt{w_{\text{IQI}}^2 + w_{\text{QQI}}^2} \quad (2.28)$$

where w_{III} , w_{QII} are the results of correlating with an in-phase replica of the BOC signal and are identical to w_{II} , w_{QI} in (2.25). The first subscript denotes mixing with an in-phase (I) or quadrature (Q) carrier replica; the second subscript denotes mixing with an in-phase or quadrature subcarrier replica, and the third subscript denotes mixing with an in-phase or quadrature (orthogonal) code replica. Figure 2.18 shows the resulting SCC BOC search correlation for a BOC $(2 \times f_C, f_C)$ signal with a sine-phased subcarrier.

Section 2.6.4 discusses mathematical representations of the w_{IQI} and w_{QQI} correlations (2.34), which can be used to enable tracking of the BOC subcarrier.

Figure 2.19 compares the PSK search correlation and equivalent BOC search correlation. The SCC technique creates a stepped correlation function that adequately approximates a single correlation peak. Including the quadrature carrier correlations w_{QII} and w_{QQI} removes the phase dependency of the search correlation, allowing location with only a coarse frequency lock. The SCC technique has been shown to deliver equivalent performance to SSB search across a range of carrier-to-noise densities (22 to 32 dB-Hz) [16]. Extensive analysis and comparison of SSB and SCC search techniques is given via Monte Carlo simulations in [17].

The hardware requirements of the SSB and SCC techniques over conventional PSK architectures are described in [1].

2.6 Tracking BOC Signals

The standard BOC early minus late discriminator curve contains multiple zero crossings, only one of which corresponds to the correct timing location (Figure

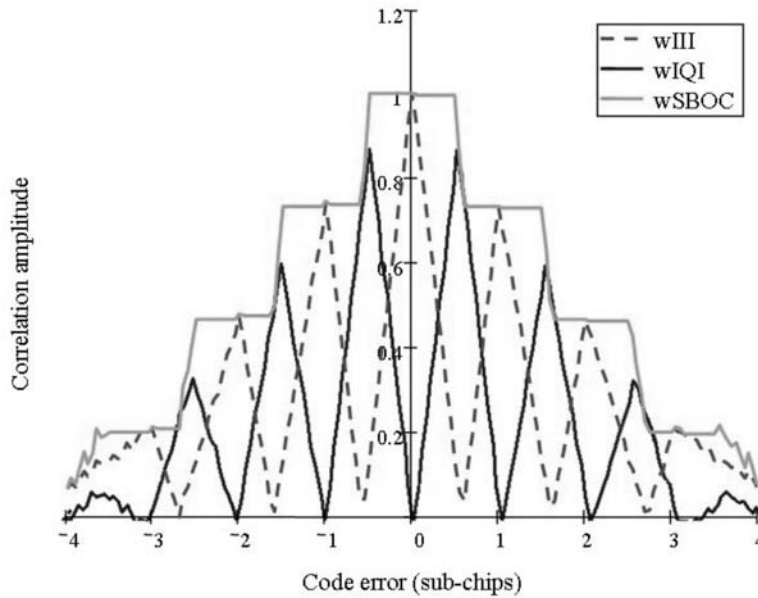


Figure 2.18 Magnitude envelopes of the w_{III} , w_{IQI} , and w_{SBOC} correlations for a BOC ($2 \times f_C$, f_C) signal.

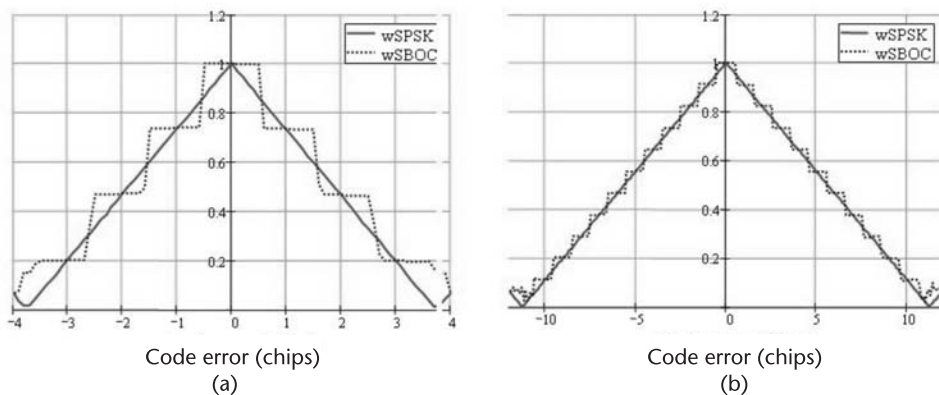


Figure 2.19 Comparison of PSK and BOC search correlations using the SCC search technique: (a) BOC ($2 \times f_C$, f_C) and (b) BOC ($6 \times f_C$, f_C).

2.20). A GNSS BOC receiver aims to achieve and maintain lock on the incoming signal at the correct timing location. In order to do so the receiver must action an algorithm or mitigation technique to achieve valid timing information, effectively removing the influence of the secondary peaks in the BOC correlation. Ideally, the receiver will achieve this with no loss of tracking sensitivity and no degradation in multipath error performance.

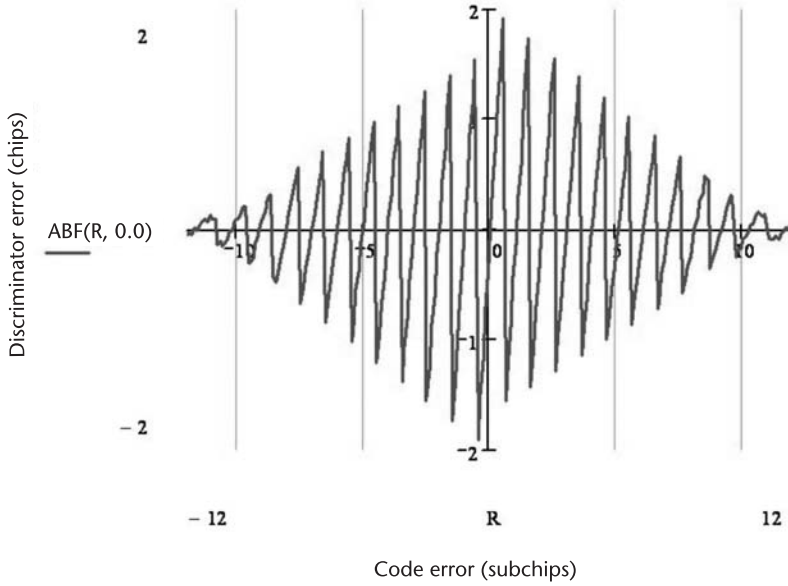


Figure 2.20 BOC ($6 \times f_C, f_C$) discriminator curve.

Here we consider four approaches for solving the ambiguity in tracking BOC signals: SSB tracking, multiple-gate discriminators (MGD), the bump-jumping (BJ) algorithm, and the dual estimator (DE).

2.6.1 BOC Tracking Using a Single Sideband (SSB)

The first approach is to use the SSB technique described in the previous section. Treating each BOC sideband as a separate PSK signal creates an unambiguous PSK correlation peak “ Λ .” Then, correlating early and late replica signals (Figure 2.21) an unambiguous discriminator curve “ $\sqrt{\Lambda}$ ” can be formed using the standard PSK discriminators (Table 2.3).

This approach provides a robust solution to resolving the BOC ambiguity. However, the root mean square (r.m.s.) timing jitter of the receiver is now dependent on the underlying PSK chipping rate and not the subcarrier rate, degrading the receiver’s timing sensitivity by a factor of $4f_S/f_C$ [18]. Therefore, this approach is only suitable for low-precision receivers.

2.6.2 BOC Tracking with Multiple-Gate Discriminators (MGD)

Another approach is to synthesize an unambiguous discriminator curve by using a combination of multiple correlator channels, which is commonly termed the MGD discriminator [19]. This technique proposes the use of K early and K late signals, T_{DC} is the early-to-late spacing, and p is an integer count from 1 to K . The resulting correlations can be written as follows [1].

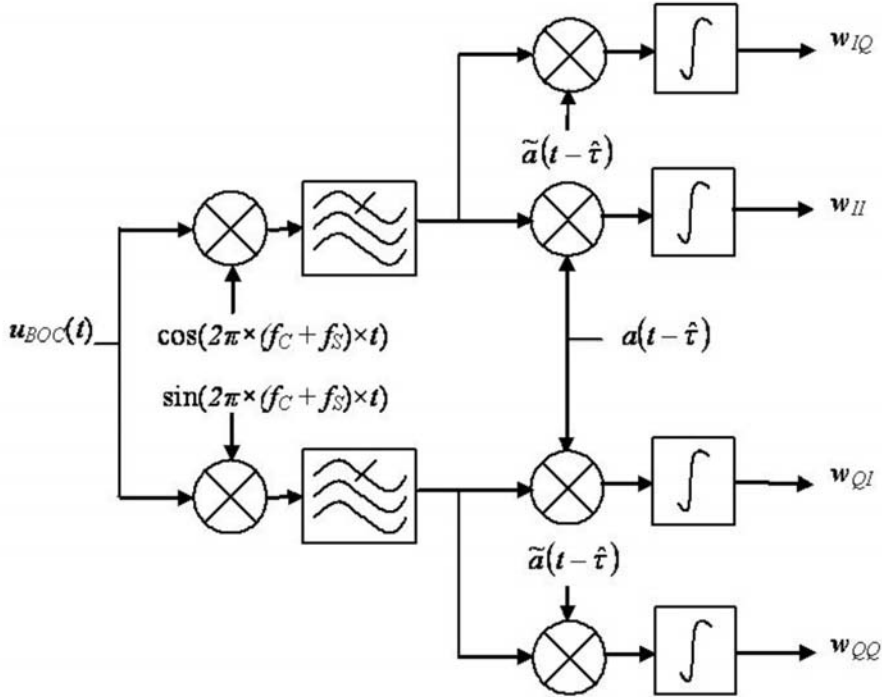


Figure 2.21 SSB BOC tracking.

$$\begin{aligned}
 \mathbf{w}_{Ip}^{(E)} &\approx A \times \cos(\hat{\phi} - \phi) \times \mathbb{W} \left(\hat{\tau} - \tau - \left(p - \frac{1}{2} \right) \times T_{DC} \right) \times d \\
 \mathbf{w}_{Ip}^{(L)} &\approx A \times \cos(\hat{\phi} - \phi) \times \mathbb{W} \left(\hat{\tau} - \tau + \left(p - \frac{1}{2} \right) \times T_{DC} \right) \times d \\
 \mathbf{w}_{Qp}^{(E)} &\approx A \times \sin(\hat{\phi} - \phi) \times \mathbb{W} \left(\hat{\tau} - \tau - \left(p - \frac{1}{2} \right) \times T_{DC} \right) \times d \\
 \mathbf{w}_{Qp}^{(L)} &\approx A \times \sin(\hat{\phi} - \phi) \times \mathbb{W} \left(\hat{\tau} - \tau + \left(p - \frac{1}{2} \right) \times T_{DC} \right) \times d
 \end{aligned} \tag{2.29}$$

A vector of coefficients can then be used to weight the influence of each of the early and late correlations in an attempt to shape an unambiguous BOC discriminator. The composite MGD discriminator is formed using of the noncoherent early-late power discriminator, whose error function can be written as

$$\begin{aligned}
 e_{\tau}(\tau) &= \sum_{p=1}^K c_p \times \left(\mathbf{w}_{Ip}^{(E)^2} + \mathbf{w}_{Qp}^{(E)^2} - \mathbf{w}_{Ip}^{(L)^2} - \mathbf{w}_{Qp}^{(L)^2} \right) \\
 e_{\tau}(\tau) &\approx \sum_{p=1}^K c_p \times \left(\mathbb{W} \left[\tau - \left(p - \frac{1}{2} \right) \times T_{DC} \right]^2 - \mathbb{W} \left[\tau + \left(p - \frac{1}{2} \right) \times T_{DC} \right]^2 \right)
 \end{aligned} \tag{2.30}$$

where c_p is the coefficient vector used to form the shape of the discriminator. Different combinations of coefficients are compared in [19] to determine the best

possible discriminator synthesis. Two classes of synthesized discriminators are defined: *smooth* and *bumpy*. *Smooth* discriminators synthesize a discriminator approaching a monotonic error function, which provides a single shallow zero crossing. *Bumpy* discriminators synthesize a single steep zero crossing with many undulations across the discriminator characteristic. Assuming a BOC ($2 \times f_C, f_C$) signal and using four early-late discriminator combinations ($K = 4$) the use of coefficients $[1 \ 1.25 \ 1.5 \ 1.75]$ and $T_{DC} = 0.525 \times T_S$ results in a smooth discriminator. Alternatively, the use of coefficients $[1 \ 1.125 \ 1.25 \ 1.375]$ and $T_{DC} = 0.2 \times T_S$ creates a bumpy discriminator. Figure 2.22 shows the composite smooth and bumpy discriminator curves.

Clearly the smooth discriminator more closely reflects the shape of a PSK “ $\sqrt{\Lambda}$ ” discriminator providing robust acquisition of the correct timing location. The bumpy discriminator has a number of nulls in the envelope, which will slow down the acquisition process and potentially cause false-lock states. It is argued that thermal noise will prevent false-lock occurring as the code error polarity is true on either side of each null. Also, it is feasible to envisage a scheme that may acquire with a smooth envelope and transition to the bumpy discriminator for precise timing location.

2.6.3 BOC Tracking with the Bump-Jumping (BJ) Algorithm

The BJ algorithm proposed in [20] determines whether or not the correct correlation peak is being tracked by comparing the amplitude of the peak currently being tracked to the amplitude of the adjacent peaks. This is achieved through the correlation of two additional offset replica codes called very early (VE) and very

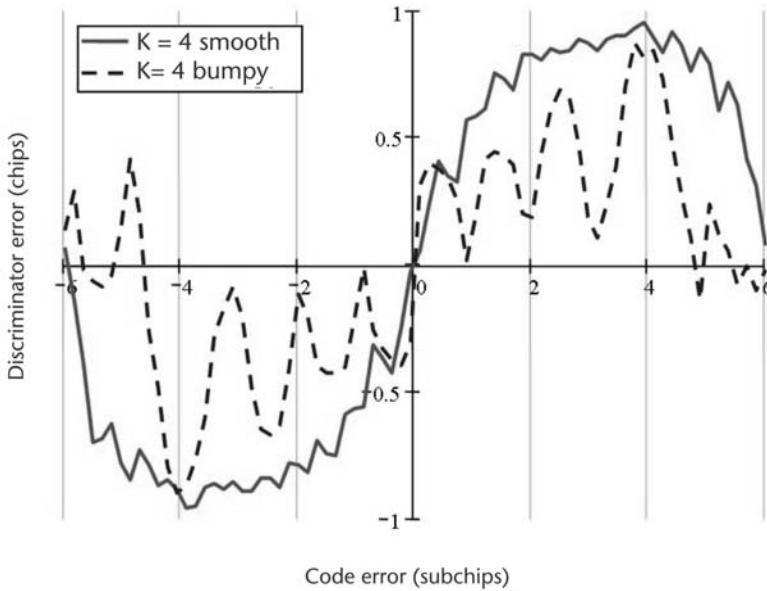


Figure 2.22 Composite MGD discriminators for a BOC ($2 \times f_C, f_C$) signal.

late (VL). These replicas are separated from the prompt (P) replica by a subchip, $\pm T_S$.

This algorithm is implemented using three counters, associated with the VE, P, or VL samples. The amplitudes of VE, P, and VL in-phase samples are compared at the end of each integrate-and-dump period. If the VE sample is larger than its counter is incremented and the VL counter decremented; if VL is larger then the opposite occurs. When the P sample is the largest, both VE and VL counters are decremented. A jump to a new peak occurs only if the VE or VL counters reach a specified threshold before the prompt counter. The counters are never decremented below zero and are reset when a threshold is reached. Figure 2.23 shows an example of a false-lock condition for a BOC ($2 \times f_C, f_C$), with the necessary tracking gates. The narrowly spaced early-late gates provide fine tracking on the signal, preserving the BOC discriminator with no sensitivity loss.

The BJ algorithm can only correct one peak at a time, shifting the tracking point by a subchip each jump and resetting all counters. Conventional search processes acquire the signal with an accuracy of $\pm T_C/2$; therefore a number of corrections (jumps) may be necessary before a valid lock is established. Assuming a search accuracy of $\pm T_C/2$, the maximum number of jumps required is equal to the ratio of subcarrier frequency to code rate f_S/f_C . Therefore, a major drawback to the BJ algorithm can be the time taken to reach the correct timing location either from acquisition or from a slip in tracking. The BJ algorithm is effectively “blind” to the number of subchips required to find the valid timing location; it must correct one peak at a time in subchip steps. In contrast, techniques providing a discriminator similar to that of a PSK signal, such as the SSB technique, the smooth MGD [19], and the DE technique can make corrections across the whole discriminator characteristic in a single step.

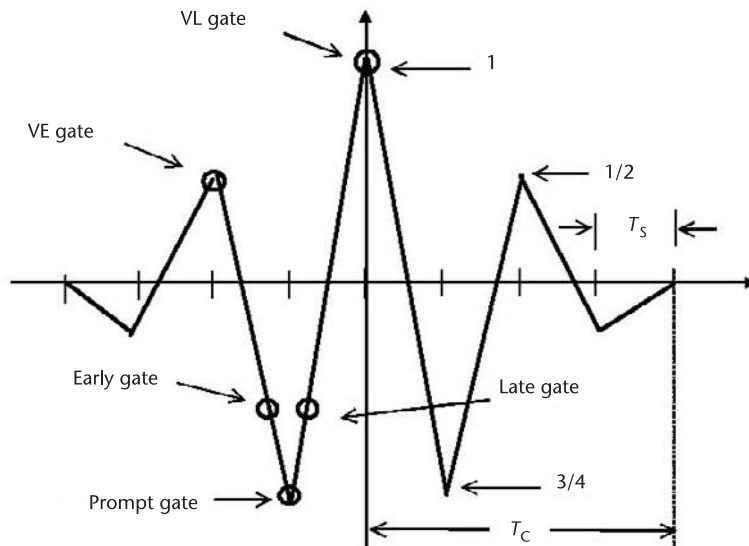


Figure 2.23 BOC ($2 \times f_C, f_C$) false-lock example with BJ gates.

2.6.4 BOC Tracking with the Dual Estimator (DE)

In a BOC transmission the subcarrier is necessarily locked to the code sequence. Indeed, it is usually seen as part of the code sequence. The time delay τ must obviously be the same in both the code and the subcarrier. However, there is nothing to stop us multiplying by component multiplicative structures with independent time-delay estimates. The DE technique is a three-loop receiver [1, 21]. The innermost DLL tracks the code phase of the received signal. The middle subcarrier locked loop (SLL) tracks the subcarrier phase of the received signal. The third loop tracks the carrier frequency and/or phase of the received signal.

The DE technique depends essentially on the fact that the subcarrier is half-periodic over a relatively short subchip width T_S and (2.24) is mathematically identical to

$$u(t) = A \times \cos(\omega_0 t + \phi) \times s(t - \tau^*) \times a(t - \tau) \times d(t)^* \quad (2.31)$$

where

$$\tau^* = \tau + nT_S \quad (2.32)$$

is a multivalued offset delay with a number of values, each offset from the delay τ by an integer n times the subchip width T_S . The equivalent polarity $d(t)^* = d(t)$ for an even number of shifts and $d(t)^* = -d(t)$ for an odd integer shift.

It should be understood that the actual subcarrier delay and the code delay for any received signal are still the same τ . The receiver must always estimate this actual nonambiguous delay τ in the code function $a(\cdot)$. It is however only necessary for the receiver to seek to estimate the ambiguous τ^* in the subcarrier function $s(\cdot)$ to achieve the same result (as if it were estimating the actual delay τ). Accordingly, the offset delay τ^* and delay τ are treated as independent quantities without regard to (2.32), and two independent estimates may be generated. Only in a final correction stage is it admitted that an estimate of offset delay τ^* and delay τ are related [as in (2.32)]. The value of estimate $\hat{\tau}$ is used to remove the ambiguities in the value of estimate $\hat{\tau}^*$. Figure 2.24 shows the general schematic of the DE receiver.

The DE BOC receiver requires three local oscillators, which generate continuous carrier, subcarrier, and code waveforms. The carrier oscillator generates an in-phase replica carrier $\cos(\omega_0 t + \hat{\phi})$ and an orthogonal replica carrier $\sin(\omega_0 t + \hat{\phi})$, where $\hat{\phi}$ is the estimate of the received carrier phase ϕ . The subcarrier oscillator generates an in-phase replica subcarrier $s(t - \hat{\tau}^*)$ and an orthogonal replica carrier $\tilde{s}(t - \hat{\tau}^*)$, where $\hat{\tau}^*$ is the SLL estimate of the time delay τ . The code oscillator generates an in-phase replica code $a(t - \hat{\tau})$ and an orthogonal replica code $\tilde{a}(t - \hat{\tau})$, where $\hat{\tau}$ is the DLL estimate of the time delay τ .

We define an orthogonal subcarrier as the difference between early and late time shifts of the subcarrier, which can be written as follows.

$$\tilde{s}(t - \hat{\tau}^*) = s\left(t - \hat{\tau}^* + \frac{T_{DS}}{2}\right) - s\left(t - \hat{\tau}^* - \frac{T_{DS}}{2}\right) \quad (2.33)$$

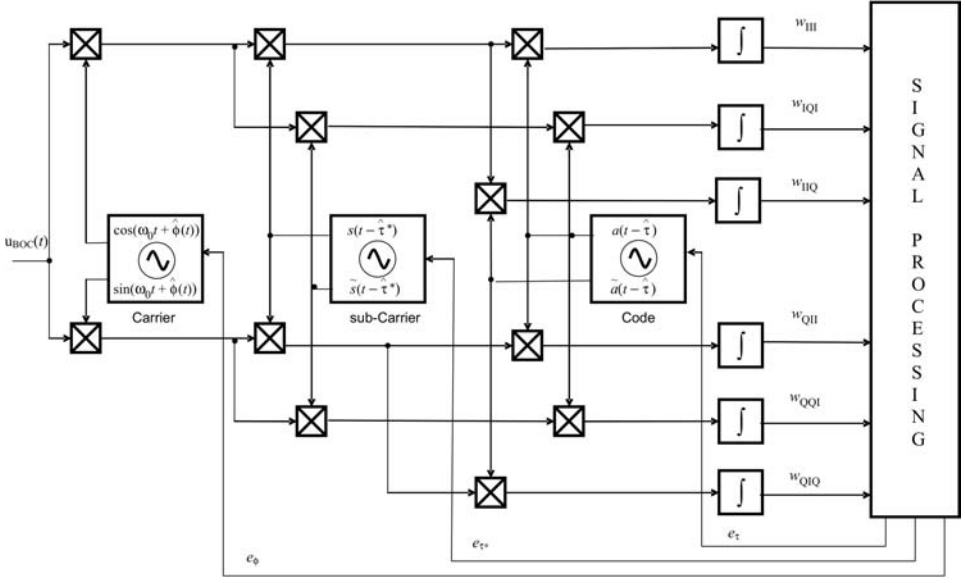


Figure 2.24 General schematic of a coherent DE BOC receiver.

T_{DS} is the total separation between “early” and “late” subcarrier waveforms, bounded by $0 < T_{DS} \leq T_S$. After mixing and integration over time T the resulting correlations can be written as follows [1].

$$\begin{aligned}
 w_{III} &\approx A \times \cos(\phi - \hat{\phi}) \times \text{trc}(\hat{\tau}^* - \tau) \times \Lambda(\hat{\tau} - \tau) \times d \\
 w_{IIQ} &\approx A \times \cos(\phi - \hat{\phi}) \times \text{trc}(\hat{\tau}^* - \tau) \times \sqrt{\Lambda}(\hat{\tau} - \tau) \times d \\
 w_{IQI} &\approx A \times \cos(\phi - \hat{\phi}) \times \text{Trs}(\hat{\tau}^* - \tau) \times \Lambda(\hat{\tau} - \tau) \times d \\
 w_{QII} &\approx A \times \sin(\phi - \hat{\phi}) \times \text{trc}(\hat{\tau}^* - \tau) \times \Lambda(\hat{\tau} - \tau) \times d \\
 w_{QOI} &\approx A \times \sin(\phi - \hat{\phi}) \times \text{trc}(\hat{\tau}^* - \tau) \times \sqrt{\Lambda}(\hat{\tau} - \tau) \times d \\
 w_{QQI} &\approx A \times \sin(\phi - \hat{\phi}) \times \text{Trs}(\hat{\tau}^* - \tau) \times \Lambda(\hat{\tau} - \tau) \times d
 \end{aligned} \tag{2.34}$$

The single-peaked $\Lambda(\cdot)$ function and the tracking $\sqrt{\Lambda}(\cdot)$ function are the same as defined for the PSK GNSS receiver in (2.9) and (2.19), respectively. The $\text{trc}(\cdot)$ function is a continuous triangular cosine waveform shown in Figure 2.16 where $\text{trc}(\hat{\tau}^* - \tau) \rightarrow \pm 1$ as $\hat{\tau}^* \rightarrow \tau + n \times T_S$. The $\text{Trs}(\cdot)$ function is a continuous trapezium shaped sine waveform where $\text{Trs}(\hat{\tau}^* - \tau) \rightarrow 0$ as $\hat{\tau}^* \rightarrow \tau + n \times T_S$, shown in Figure 2.25. The $\text{Trs}(\cdot)$ function is identical to the result of early minus late correlations of the $\text{trc}(\cdot)$ function as follows.

$$\text{Trs}(\hat{\tau}^* - \tau) = \text{trc}\left(\hat{\tau}^* - \tau - \frac{T_{DS}}{2}\right) - \text{trc}\left(\hat{\tau}^* - \tau + \frac{T_{DS}}{2}\right) \tag{2.35}$$

Discriminators for FLL and PLL implementations of the DE can be found in [1].

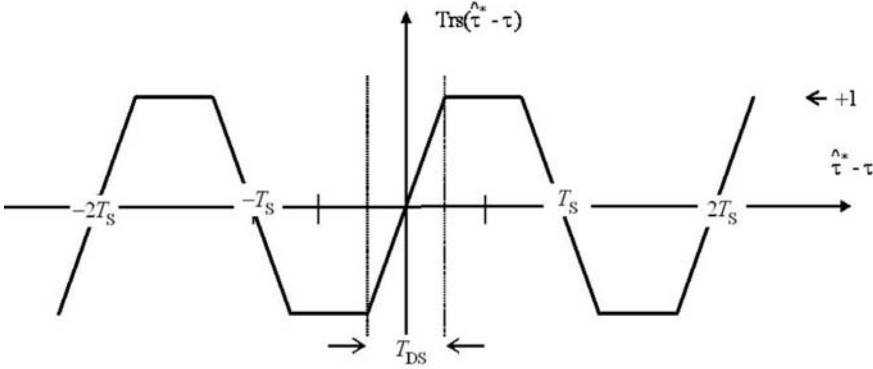


Figure 2.25 Trapezoidal sine function.

An example acquisition of the DE BOC receiver is shown in Figure 2.26 for a BOC($2 \times f_C, f_C$) signal, derived from Mathcad simulation with equal DLL and SLL loop bandwidths ($B_{DLL} = B_{SLL} = 1$ Hz) and carrier to noise density, $C/N_0 = 30$ dB-Hz. The initial delay offset was set at 2.5 subchips and the initial phase error at $\pi/4$.

Under loop operations the DLL delay estimate provides unambiguous tracking with timing jitter equivalent to that of the underlying PSK modulation. The SLL delay estimate $\hat{\tau}^*$ delivers the full tracking accuracy of the BOC modulation [1]; it is however ambiguous, locking to integer subchip values. Once the loops have

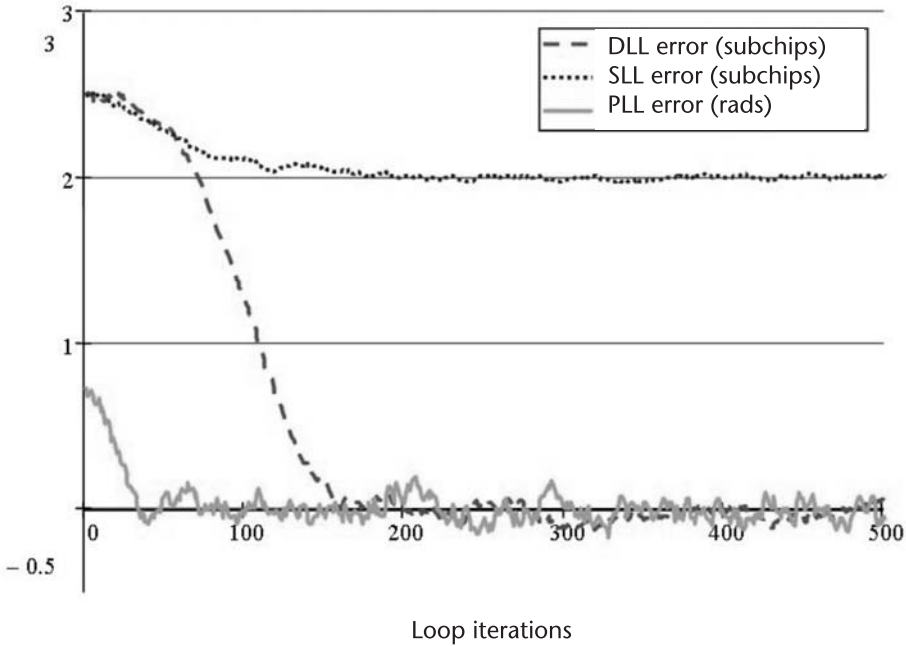


Figure 2.26 Example acquisition of the DE BOC receiver ($B_{DLL} = B_{SLL} = 1$ Hz, $C/N_0 = 30$ dB-Hz).

settled the ambiguity of the SLL estimate can be resolved through the noisier but unambiguous DLL estimate as follows.

$$\hat{\tau}^+ = \hat{\tau}^* - \text{round}\left(\frac{\hat{\tau}^* - \hat{\tau}}{T_S}\right) \times T_S \quad (2.36)$$

where $\hat{\tau}^+$ is the corrected delay estimate. Figure 2.27 shows an example acquisition with the corrected delay estimate. The DE technique can also be extended to track AltBOC modulation as shown in [1] and MBOC modulation as shown in [22].

A key performance measure of particular interest is that of the multipath performance of the DE BOC receiver. The simplest method of evaluating the multipath error performance is to consider the effect of a single interfering multipath signal with various relative time delays. This provides only a worse case analysis of error due to multipath but does provide a measure with which we can compare BOC tracking schemes. The result of this analysis is a multipath error envelope that, for conventional receivers, is dependent on the correlation function and the type of discriminator implemented for code tracking. Therefore, to provide a fair comparative basis on which to evaluate multipath error envelopes it is important to set equal discriminator spacing.

Figure 2.28 shows the multipath error envelopes of a conventional (early-minus-late discriminator) BOC receiver and the DE BOC receiver for a BOC(2,1) signal. The DE BOC error envelope is computed by analyzing the error of the corrected SLL delay estimate with multipath interference. The relative performance of each scheme can be seen by computing the running average error across the

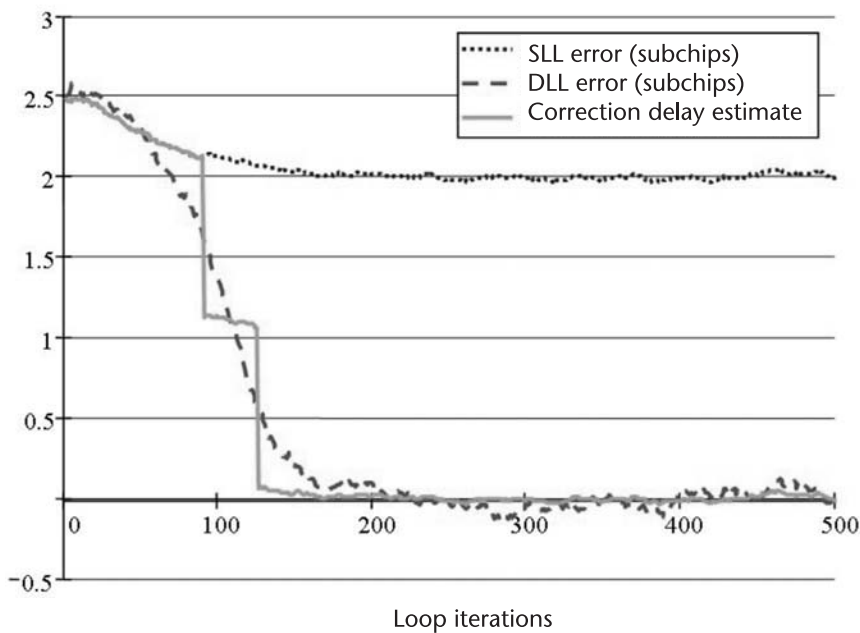


Figure 2.27 Corrected delay estimate of the DE BOC receiver ($B_{DLL} = B_{SLL} = 1$ Hz, $C/N_0 = 30$ dB-Hz).

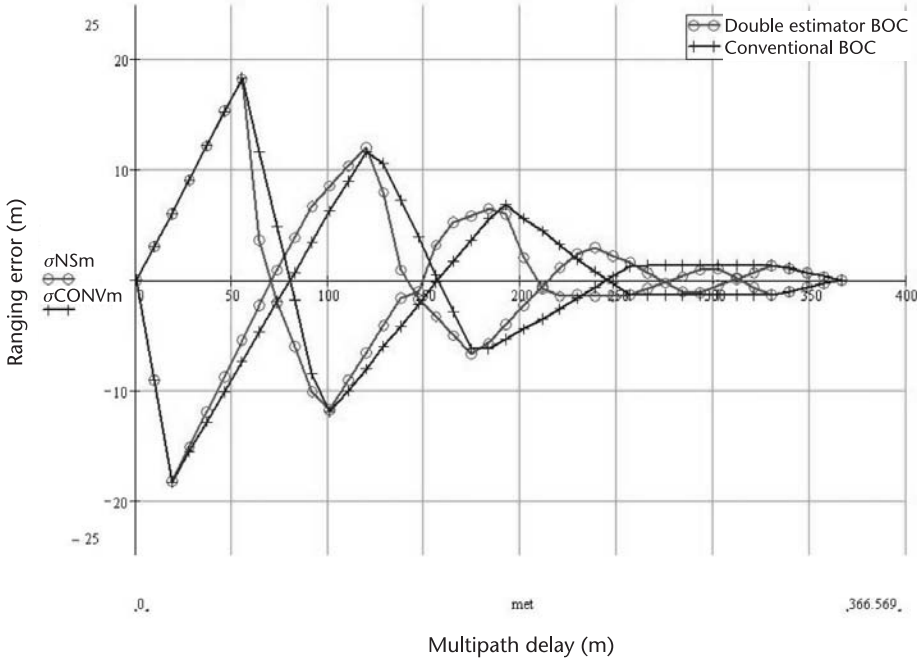


Figure 2.28 Multipath error envelope of a conventional and DE BOC receiver, BOC (2,1) $\Delta_{BOC} = T_S$.

data set. The pattern of the DE BOC multipath envelope broadly follows that of the conventional receiver but does show a slight improvement (8.2%) across the whole dataset.

A comparison of the BOC tracking techniques described in this chapter is given in [1], where receiver timing jitter, multipath performance, hardware impact, and reliability are considered. From this analysis, a design trade-off is identified between receiver settling time and timing jitter of the delay estimate in all techniques except the DE. Techniques approximating a PSK single correlation peak, such as SSB and smooth MGD, provide an unambiguous rapid loop settling time at the cost of significantly worse receiver timing jitter. Other techniques, such as the BJ algorithm and bumpy MGD, can provide apparent unambiguous BOC tracking with little or no sensitivity loss but at the cost of increasing acquisition and slip correction time. In [23] the reliability of the BJ algorithm is shown to be particularly sensitive to noise, signal distortion, and multipath, especially for BOC signals with a high ratio of subcarrier-to-code rate.

The DE BOC receiver is believed to provide the best of both worlds: an unambiguous rapid loop settling time with the full timing precision and multipath error performance available from the BOC characteristic. The settling time of the DLL follows the desirable single correlation peak synonymous with PSK modulation, while the SLL simply settles to the nearest subchip offset. The maximum precision of the BOC modulation is then achieved by correcting the SLL estimate using the DLL estimate as in (2.36). Tracking slips or disturbances in the subcarrier loop

caused by noise, multipath, or distortion are then automatically corrected by the unambiguous code loop.

References

- [1] Blunt, P. D, *Advanced Global Navigation Satellite System Receiver Design*, Ph.D. thesis, University of Surrey, U.K., May 2007.
- [2] Sklar, B., *Digital Communications: Fundamentals and Applications*, Upper Saddle River, NJ: Prentice Hall International, 1988, pp. 695–757.
- [3] GPS Navstar interface control documents, available at: <http://www.navcen.uscg.gov/gps/geninfo/>, accessed in May 2007.
- [4] Betz J. W., “The Offset Carrier Modulation for GPS Modernisation,” *Proceedings of ION 1999 National Technical Meeting*, San Diego, California, January 1999.
- [5] European Space Agency, Galileo Joint Undertaking, “Galileo Open Service Signal In Space Interface Control Document” Draft 0, available at: <http://www.galileoju.com>, accessed in May 2006.
- [6] Hein G. W., Avila-Rodríguez J.-A., et al, “MBOC: The New Optimized Spreading Modulation Recommended for GALILEO L1 OS and GPS L1C,” *Proceedings of IEEE/ION PLANS 2006*, San Diego, California, April 2006.
- [7] Weiler, R. M., et al, “Performance of an L1/E5 GNSS Receiver Using a Direct Conversion Front-End Architecture,” *Proceedings of ION GPS 2008*, Savannah, Georgia, September 2008.
- [8] Unwin, M. J., *The Design and Implementation of a Small Satellite Navigation Unit Based on a Global Positioning System Receiver*, Ph.D. dissertation, University of Surrey, 1995.
- [9] Mitel Semiconductor, “GP2000 GPS Chipset—Designers Guide,” available at: <http://www.datasheetarchive.com/>, accessed in December 2006.
- [10] Van Nee, D. J. R., and A. J. R. M. Coenen, “New Fast GPS Code-Acquisition Technique using FFT,” *Electronics Letters*, Vol. 27, January 1991.
- [11] Kilvington, J., “Receivers for Navigation Satellite Systems,” U.S. Patent 4601005, July 1986.
- [12] Ward, P. W., J. W. Betz, and C. J. Hegarty, “Satellite Signal Acquisition, Tracking and Data Demodulation,” in *Understanding GPS Principles and Applications* (Second Edition), edited by E.D. Kaplan and C.J. Hegarty, Norwood, MA: Artech House, 2006, pp. 153–241.
- [13] Groves, P. D., “Satellite Navigation Processing, Errors, and Geometry,” in *Principles of GNSS, Inertial and Multisensor Integrated Navigation Systems*, Norwood, MA: Artech House, 2008, pp. 195–277.
- [14] Martin, N., et al, “BOC(x,y) Signal Acquisition Techniques and Performances,” *Proceedings of ION GPS 2003*, Portland, Oregon, September 2003.
- [15] Ward, P.W., “A Design Technique To Remove the Correlation Ambiguity in Binary Offset Carrier (BOC) Spread Spectrum Signals,” *ION 59th Annual Meeting/CIGTF 22nd Guidance Test Symposium*, Albuquerque, New Mexico, June 2003.
- [16] Heiries, V., et al, “Analysis of Nonambiguous BOC Signal Acquisition Performance,” *Proceedings of ION GNSS 2004*, Long Beach, California, September 2004.
- [17] Weiler, R., et al, “The Effect of Cosine-Phased BOC Modulation on the GNSS Receiver Search Process,” *The Journal of Navigation*, Volume 61, Number 4, October 2008.
- [18] Bello, P. A., and R. L. Fante, “Code Tracking Performance for Novel Unambiguous M-Code Time Discriminators,” *Proceedings of ION National Technical Meeting*, San Diego, California, January 2005.

- [19] Fante, R. L., “Unambiguous Tracker for GPS Binary-Offset-Carrier Signals,” *ION 59th Annual Meeting/CIGTF 22nd Guidance Test Symposium*, Albuquerque, New Mexico, June 2003.
- [20] Fine, P., and W. Wilson, “Tracking Algorithm for GPS Offset Carrier Signals,” *Proceedings of ION 1999 National Technical Meeting*, San Diego, CA, January 1999.
- [21] Hodgart M. S, P. D. Blunt, and M. Unwin, “The Optimal Dual Estimate Solution for Robust Tracking of Binary Offset Carrier (BOC) Modulated GNSS,” *ION-GNSS*, Fort Worth, September 2007.
- [22] Hodgart, M. S., R. Weiler, and M. Unwin, “A Triple Estimating Receiver of Multiplexed Binary Offset Carrier MBOC Modulated Signals,” *ION-GNSS*, Savannah, Georgia, September 2008.
- [23] Blunt, P. D., et al, “Demonstration of BOC(15, 2.5) Acquisition and Tracking with a Prototype Hardware Receiver,” *European Navigation Conference*, Geneva, May 2007.

GNSS Navigation: Estimating Position, Velocity, and Time

Scott Gleason and Demoz Gebre-Egziabher

3.1 Overview

This chapter demonstrates how one generates estimates of position, velocity, and time (PVT) using GNSS. The standard least-squares estimation method is used to demonstrate this. The algorithms presented in this chapter are not new and have been discussed extensively in the GNSS literature [1]. Our intent is to give a hands-on demonstration of performing these calculations. This is motivated by our observation that the existing GNSS literature has often lacked representative examples demonstrating the end-to-end process of generating a PVT solution. The best, and perhaps most entertaining, way to achieve this goal is to provide a hands-on experience through use of a simulator that can be configured to the readers' specifications. The simulator allows readers to work with simulated measurements collected from a receiver placed at a location and time of their choosing. The receiver could be running on the roof of one's house on his or her birthday, for example. The simulation environment has been designed to allow readers to experiment with placing and moving simulated receivers around the Earth under modeled GNSS constellations and to reproduce realistic and changing solution geometries. Our hope is that with a little effort readers can get a better feel for the GNSS PVT estimation problem and appreciate some of the subtleties of the calculations under different conditions and constraints.

The simulator presented with the book incorporates features that are not yet fully available to actual GNSS users. For example, it has the ability to simulate fully operational Galileo and GLONASS constellations. It can also simulate any generic GNSS constellation in that it allows for up to 100 satellites in any orbit to be configured as GNSS transmitters. Furthermore, the simulator can easily provide dual-frequency measurements for illustrating the improved performance possible using multifrequency measurements.

It should be noted that, while the data generated by the simulator is designed to be as realistic as possible, it is intended to be a pedagogical tool for demonstrating how to perform the basic PVT calculations. It is provided for readers' education and, perhaps, amusement. It should *not* be used as a tool for analyzing the performance of GNSSs that will be used in products that can be fielded or used outside of an educational setting.

3.2 Position, Velocity, and Time (PVT) Estimation

The primary application of a GNSS is, as the name suggests, to navigate globally. This includes using raw range measurements to estimate the PVT at a user receiver. The user receiver could be carried by someone hiking in the forest, in an automobile, on an aircraft or even bolted onto a satellite where the actual human users are the operators on the ground.

Much of this book is dedicated to how to best apply the basic PVT information to a useful end. In many cases the standard PVT estimate generated by a standard GNSS receiver is perfectly adequate for this. This is primarily the case when the user receiver has a clear unobstructed view of the sky and can easily make measurements from four or more satellites. We will show by example how as these ideal conditions degrade our knowledge of where the receiver is, how fast it is moving, and what time it is can become significantly corrupted. Nonetheless, the basic estimation steps described below form one of the basic building blocks of most navigation applications. Thus, this book contains additional chapters dealing with generating a PVT solution in challenging environments where the stand-alone estimates may not be sufficient.

3.2.1 Estimating Receiver Position and Clock Bias

The starting point for determining the position of a GNSS receiver is the pseudorange measurement, which, as a function of time, is given by the well-known equation from Misra and Enge [1]

$$\rho^k(t) = r^k(t, t - \tau) + c[\delta t_u(t) - \delta t^k(t - \tau)] + I^k(t) + T^k(t) + \epsilon^k(t) \quad (3.1)$$

where:

k is the individual identifier for the transmitting satellite.

$\rho^k(t)$ is the pseudorange measurement for satellite k (meters).

t is the signal reception time (seconds).

$r^k(t, t - \tau)$ is the true or geometric range between the receiver at the time of reception, t , and the transmitting satellite k at the time of transmission, $t - \tau$ (meters).

τ is the time of flight (sometimes called delay) for the signal during its travel between the satellite and receiver (seconds).

c is the speed of light in a vacuum (meters/second).

$\delta t_u(t)$ is the receiver clock offset relative to GPS time at the signal reception time (meters).

$\delta t^k(t - \tau)$ is the k th satellite's clock offset relative to GPS time at the time of transmission (meters).

$I^k(t)$ is the propagation delay caused by the Earth's ionosphere for satellite k (meters).

$T^k(t)$ is the propagation delay caused by the Earth's troposphere for satellite k (meters).

$\epsilon^k(t)$ is the miscellaneous unmodeled range error for satellite k . The miscellaneous range error includes, for example, uncertainty in the satellite position and clock offset as well as receiver noise and multipath.

All of the terms on the right-hand side of (3.1) except for δt_u have the superscript k . This means that, at any given time, these terms can be different for each satellite. The term δt_u , however, is common to all satellite measurements—it is a common mode error on all measurements.

With this fact in mind, (3.1) can be simplified by eliminating the specific references to time and grouping all errors into a single term called $\tilde{\epsilon}_T^k$ (subscript T for the total pseudorange error). Hence, the corrected pseudorange measurements from individual satellites can then be simplified to,

$$\rho_c^k = r^k + c\delta t_u + \tilde{\epsilon}_T^k \quad (3.2)$$

where the unknowns are the receiver clock offset δt_u and the true range r^k . The magnitude and statistical nature of $\tilde{\epsilon}_T^k$ affects the quality of the PVT solution. The smaller and more uncorrelated it is, the better the quality of the solution. Thus, it is normal to preprocess the GPS pseudorange measurements to reduce the magnitude of $\tilde{\epsilon}_T^k$ and make it as uncorrelated as possible. GNSS satellites broadcast a message that allows the user to estimate δt^k , I^k and T^k and, subsequently, remove them from $\tilde{\epsilon}_T^k$. So, if it is assumed that the satellite clock, the ionosphere, and troposphere corrections have been applied, then $\tilde{\epsilon}_T^k$ will consist of residuals from these corrections and other unmodeled errors. The incorporation of these corrections is indicated by the subscript c on the range measurements.

Using Cartesian coordinates for the user-receiver location at the signal reception time $\mathbf{x}_u = [x, y, z]$ and satellite positions at the time of transmission $\mathbf{x}^k = (x^k, y^k, z^k)$ the true (or geometric) range can be expressed as,

$$r^k = \sqrt{(x^k - x)^2 + (y^k - y)^2 + (z^k - z)^2} = \|\mathbf{x}^k - \mathbf{x}_u\| \quad (3.3)$$

To perform this calculation we require that both the position of the user receiver and the positions of the satellites are expressed in the same coordinate frame. The most common reference frame for GPS applications is WGS84, or the World Geodetic Survey 1984, Earth-centered Earth-fixed coordinate frame. Notably the WGS84 standard also defines the surface geoid that is commonly used as a local Earth surface reference. More information on Earth reference frames, including WGS84, can be found in Chapter 3 of [1] or Chapter 2 of [2]. Additionally, the paper “Modern Geodetic Reference Frames for Precise Satellite Positioning and Navigation” by Kouba and Popelar has been included on the DVD that accompanies this book and provides a useful summary of the topic. Parameters describing the precise orbits of each GPS satellite are broadcast in the navigation data message modulated onto the signal in the WGS84 reference frame. This information is used within the receiver to calculate the satellite position vectors at the time of

transmission, \mathbf{x}^k . Substituting this into (3.2) and simplifying the total receiver clock error term to $b = c\delta t_u$ gives,

$$\rho_c^k = \|\mathbf{x}^k - \mathbf{x}_u\| + b + \tilde{\epsilon}_T^k \quad (3.4)$$

Equation (3.4) relates the corrected pseudorange measurements for the k th satellite in view to the WGS84 position coordinates for the user's position vector $\mathbf{x}_u = [x, y, z]$, the user receiver clock offset b , and the pseudorange error term $\tilde{\epsilon}_T^k$. The error term can be minimized to some extent using various techniques but never completely eliminated. Section 3.2.5 briefly summarizes errors. The unknowns in this equation are the three-user position coordinates and the clock bias. Thus, given a minimum of four pseudorange measurements, its possible to estimate the four unknowns. Figure 3.1 is a visual representation of these measurements.

The standard approach for estimating the receiver position and clock offset is first to linearize the pseudorange measurements (3.4) around a rough guess of the receiver position and clock bias and then to iterate until the difference between the guess and the measurements approaches zero. While this implies that some information is needed about the initial receiver position, it turns out that the solution is not very sensitive to this initial (or rough) guess. For example, the fastgps

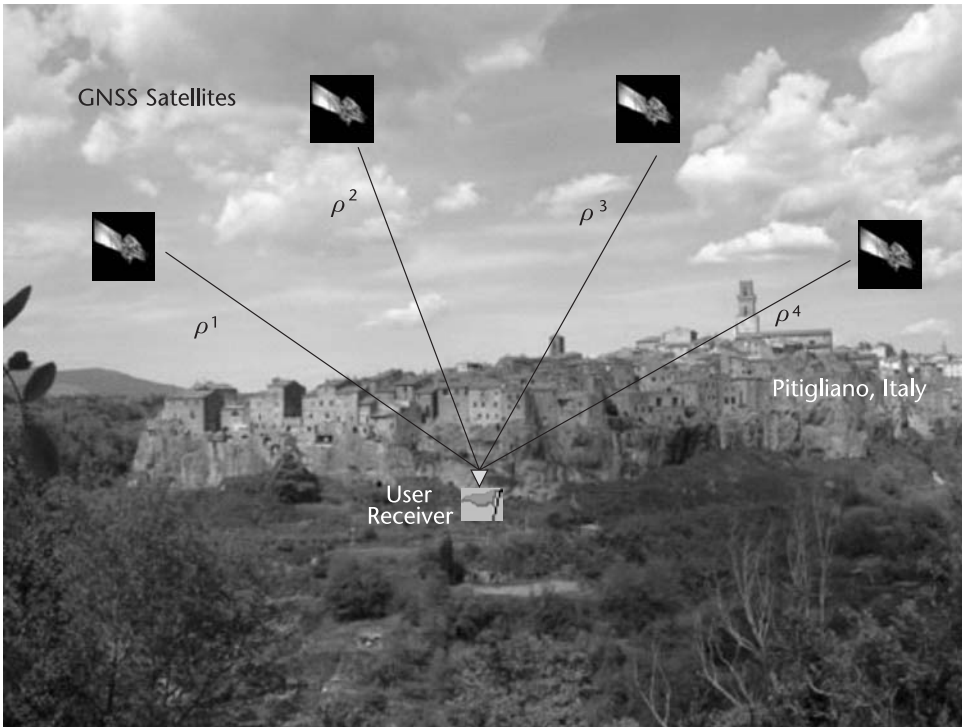


Figure 3.1 Illustration of pseudorange measurements for four GNSS satellites. A minimum of four measurements are needed to estimate the user receiver position and clock offset. In most cases more than four measurements are available.

software receiver discussed in Chapter 5 uses the center of the Earth as its initial guess. While this will work for terrestrial users, it may lead to a false solution for space-borne applications such as those described in Chapter 13. Thus, some judgment is required in selecting the initial conditions for the PVT solution. It is left as an exercise for the reader, using the examples provided, to experiment with this sensitivity. If the initial rough guess for the receiver position and clock bias are expressed as $\mathbf{x}_0 = [x_0, y_0, z_0]$ and b_0 , respectively, then the corresponding estimated pseudorange for the k th satellite is written as:

$$\rho_0^k = \|\mathbf{x}^k - \mathbf{x}_0\| + b_0 \quad (3.5)$$

Next, the estimated pseudoranges based on our initial guess ρ_0^k and the measured but corrected pseudoranges ρ_c^k from (3.4) are related using a set of linear equations. This can be done by expressing the true receiver position $\mathbf{x}_u = [x, y, z]$, as $\mathbf{x}_u = \mathbf{x}_0 + \delta\mathbf{x}$ and the true receiver clock offset as $b = b_0 + \delta b$. That is, the true receiver position and clock offset are assumed to be equal to the rough guess with a linear correction applied (a correction that could initially be quite large). Using these relationships, it is possible to express (in terms of a pseudorange correction) the estimated and corrected pseudorange expressions from (3.4) and (3.5) and relate this to the four unknowns of interest. The difference between the estimates and measured pseudoranges provides the starting point,

$$\delta\rho^k = \rho_c^k - \rho_0^k \quad (3.6a)$$

$$\delta\rho^k = \|\mathbf{x}^k - \mathbf{x}_u\| + b + \tilde{\epsilon}_T^k - [\|\mathbf{x}^k - \mathbf{x}_0\| + b_0] \quad (3.6b)$$

$$\delta\rho^k = \|\mathbf{x}^k - \mathbf{x}_0 - \delta\mathbf{x}\| + b_0 + \delta b + \tilde{\epsilon}_T^k - \|\mathbf{x}^k - \mathbf{x}_0\| - b_0 \quad (3.6c)$$

Then applying a Taylor series expansion to arrive at

$$\delta\rho^k = -\frac{(\mathbf{x}^k - \mathbf{x}_0)}{\|\mathbf{x}^k - \mathbf{x}_0\|} \cdot \delta\mathbf{x} + \delta b + \tilde{\epsilon}_T^k = -L_{unit}^k \cdot \delta\mathbf{x} + \delta b + \tilde{\epsilon}_T^k \quad (3.6d)$$

where

$$L_{unit}^k = \frac{(\mathbf{x}^k - \mathbf{x}_0)}{\|\mathbf{x}^k - \mathbf{x}_0\|} \quad (3.6e)$$

Note that L_{unit}^k is the line-of-sight unit vector between the estimated receiver location and satellite k . When measurements from more than one satellite are available, the equation above can be written compactly as follows:

$$\delta \boldsymbol{\rho} = \begin{bmatrix} \delta \rho^1 \\ \delta \rho^2 \\ \dots \\ \delta \rho^K \end{bmatrix} = \begin{bmatrix} (-\mathbf{L}_{unit}^1)^T & 1 \\ (-\mathbf{L}_{unit}^2)^T & 1 \\ \dots & 1 \\ (-\mathbf{L}_{unit}^K)^T & 1 \end{bmatrix} \begin{bmatrix} \delta \mathbf{x} \\ \delta b \end{bmatrix} + \begin{bmatrix} \tilde{\epsilon}_T^1 \\ \tilde{\epsilon}_T^2 \\ \dots \\ \tilde{\epsilon}_T^K \end{bmatrix} \quad (3.7)$$

where K is the total number of satellite measurements generated by the receiver and must be greater than or equal to four to solve for all four unknowns. A superscript T indicates a matrix transpose operation, as opposed to the subscript T or “total” error symbol (epsilon symbol). Equation (3.7) can be written in terms of what is often called the geometry matrix, G , as follows:

$$\delta \boldsymbol{\rho} = G \begin{bmatrix} \delta \mathbf{x} \\ \delta b \end{bmatrix} + \tilde{\epsilon}_T \quad (3.8a)$$

where,

$$G = \begin{bmatrix} (-\mathbf{L}_{unit}^1)^T & 1 \\ (-\mathbf{L}_{unit}^2)^T & 1 \\ \dots & 1 \\ (-\mathbf{L}_{unit}^K)^T & 1 \end{bmatrix} \quad (3.8b)$$

The line-of sight vectors within the geometry matrix collectively determine the overall quality of the position estimation geometry. As discussed in detail in Section 3.2.3, the matrix is used to calculate the dilution of precision (DOP) values, which give a good indication of the estimated position accuracy.

Before these equations are solved for our guess correction values, $\delta \mathbf{x}$ and δb , it is important to discuss a few subtle issues associated with (3.7). The first one is that the convergence of the iterative solution will depend on the geometry of the receiver-satellites system, which, in turn, affects the rank of G . Problems can occur if G is rank-deficient or close to it, which can occur when all the satellites lie in or very close to the same plane in three-dimensional space. As an example, imagine a receiver traveling north through an urban canyon of tall buildings. Imagine that the buildings lie on the east and west side of the direction of travel. Navigating in the midst of such tall buildings will result in the visibility of the sky (at the receiver) to be good in one plane (north and south), but poor in the perpendicular direction (east and west), where buildings block a significant number of satellites that would normally provide measurements. In conditions such as these it is conceivable that the satellites tracked may come close to lying in a single plane and, thus, create a condition where G is rank-deficient or close to it. This can severely degrade the position estimate.

A second subtlety is associated with the errors that can result if the reference frame rotation during the propagation delay τ is not corrected. This is the small

angular displacement (at a rate of approximately 72 microradians per second) made during the time it took for the signal to travel from the satellite to the receiver. Compensation for this motion can be made by rotating the original estimate for the satellite position vector $\tilde{\mathbf{x}}^k$ backwards by a small amount to obtain the correct value of \mathbf{x}^k . This is done using the approximate transmission delay τ and the rotation rate of the Earth ω_E , such that the value for the satellite position is,

$$\mathbf{x}^k = \begin{bmatrix} \cos(\omega_E \tau) & \sin(\omega_E \tau) & 0 \\ -\sin(\omega_E \tau) & \cos(\omega_E \tau) & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \tilde{\mathbf{x}}^k \quad (3.9)$$

Ideally, this correction should be iterated as the value τ changes with the satellite position correction. However, in practice a single correction is almost always sufficient to reduce the error to acceptable levels.

If the errors in the vector $\tilde{\mathbf{e}}_T$ are zero-mean and uncorrelated, then the optimal estimate for $\delta\mathbf{x}$ and δb in (3.7) (in the least squares sense) would be given by:

$$\begin{bmatrix} \delta\mathbf{x} \\ \delta b \end{bmatrix} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \delta\boldsymbol{\rho} \quad (3.10)$$

Note that this equation performs a least squares optimization only when the solution is over-determined (i.e., $k > 4$). When there are only the minimum four measurements, the result is the solution of a set of linear equations. Regardless, solving this equation gives us corrections for our initial guess, $\delta\mathbf{x}$ and δb , which can now be reapplied to the initial guess, and the whole process repeated. The new initial guess for the next iteration is updated such that,

$$\mathbf{x}_{0|\text{new}} = \mathbf{x}_{0|\text{old}} + \delta\mathbf{x} \quad (3.11a)$$

$$b = \delta b \quad (3.11b)$$

The above steps [starting at (3.5)] are then repeated until the correction vectors converge towards zero.

At this point, there are two important points that should be highlighted. First, when (3.5) is calculated the value of b_0 is normally zero. This is due to the fact that the clock bias is calculated on every iteration based on the estimated position. Therefore, when the position solution has converged the value of the clock bias at this iteration δb will be the best estimate. This process is illustrated in Figure 3.2 as well as demonstrated using simulated data input into MATLAB/Octave (GNSS_Sim_PVT.m) and Python (GNSS_Sim_PVT.py) scripts.

Second, when generating the least squares estimate of $\delta\mathbf{x}$ and δb via (3.10) it was assumed that all the pseudorange measurements used in the procedure are of equal quality, which is almost never the case. It is often the case that satellites tracked at lower elevations (or rising or setting satellites) contain greater atmospheric errors and can make the position estimate worse rather than better. These issues as well

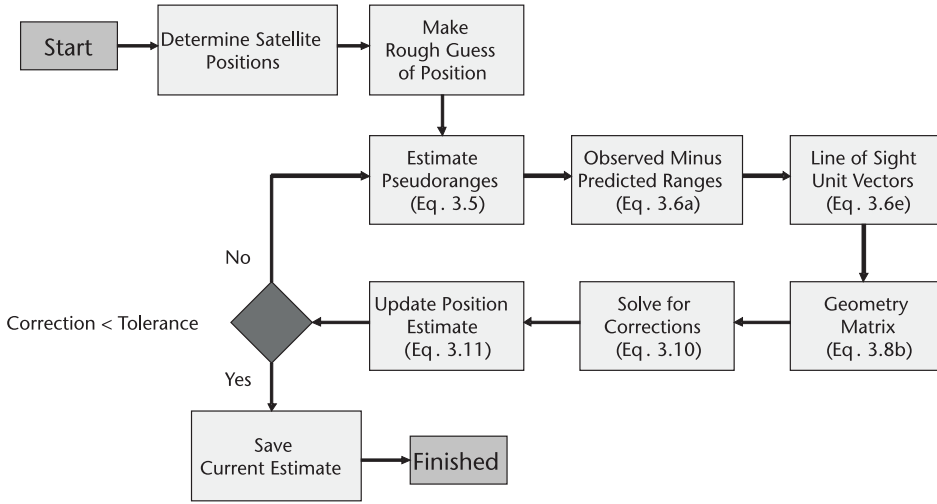


Figure 3.2 Illustration of calculation steps in the iterative position calculation.

as ways to quantify and eliminate such situations are discussed in greater detail in GNSS textbooks such as Misra and Enge [1] or Kaplan and Heggarty [2]. The reader is encouraged to consult these texts for further information.

3.2.2 Impact of Ionosphere Errors

Ionosphere errors can be large, especially during times of high solar activity. In many cases, this error can be large enough to significantly limit the effectiveness of GNSS in certain applications. To get a better idea as to the magnitude of typical ionospheric errors (where a pseudorange error magnitude of greater than 5m is not uncommon) please refer to the well-known model of Klobuchar [3], which is applied during the simulations described below. Fortunately, in the near future GNSS signals on multiple frequencies will be available to civilian users (e.g., GPS L2 and L5). By combining measurements made at different frequencies, the pseudorange errors due to the ionosphere on each frequency (that is, I^k) can be estimated. Thus, estimated values of I^k can be subtracted from (3.1) to remove the ionosphere error.

Consider the following two measurements taken from satellite k : $\rho_{f_1}^k$ and $\rho_{f_2}^k$, where f_1 and f_2 are the frequencies of the two pseudorange measurements. These could be any combination of frequencies transmitted from the same satellite that will soon be available. These two pseudorange measurements can be used to directly observe or estimate the ionosphere-induced range error and, thereby, improve the pseudorange measurement. The delay due to the ionosphere at a single frequency, say f_1 , can be estimated using the following equation taken from [1]:

$$I_{f_1}^k = \frac{f_2^2}{(f_1^2 - f_2^2)} (\rho_{f_2}^k - \rho_{f_1}^k) \quad (3.12)$$

For the case of GPS where $f_1 = \text{L1}$ and $f_2 = \text{L2}$, this becomes:

$$I_{f_1}^k = 1.546 \times (\rho_{f_2}^k - \rho_{f_1}^k) \quad (3.13)$$

Note that, as discussed in [1] these measurements do not correct for other system errors (such as those added by the troposphere). Furthermore, the process of combining two noisy pseudorange measurements can result in a factor of three increase in the measurement noise. However, given the often large magnitude of the ionospheric error being eliminated the net gain of this technique is almost always positive with regards to positioning accuracy.

3.2.3 Impact of Satellite-User Geometry (DOP)

The satellite-user geometry can have a large impact on the accuracy of the PVT estimates obtained from GNSSs. In other words, some satellite-user geometries will result in a higher accuracy solution than others. As such, it is useful to have a way of comparing different satellite-user geometries. The metric normally used for measuring this impact is dilution of precision (DOP), which represents the degree to which satellite-user geometry dilutes the accuracy of the PVT. By analogy, if your position were represented by a glass of high-quality Scotch whiskey, the DOP would be a measure of the amount of added water. The higher the DOP, or the more water added, the lower the quality of your position estimate or the enjoyment of your drink, whichever the case may be.

As previously mentioned, information about DOP is encoded in the geometry matrix G . The DOP matrix H is computed from the geometry matrix as follows:

$$\mathbf{H} = (\mathbf{G}^T \mathbf{G})^{-1} = \begin{bmatrix} H_{11} & — & — & — \\ — & H_{22} & — & — \\ — & — & H_{33} & — \\ — & — & — & H_{44} \end{bmatrix} \quad (3.14)$$

If the satellite and user positions are expressed in East-North-Up (ENU) coordinates, then the square root of the diagonal entries of the H matrix are called the east DOP $EDOP = \sqrt{H_{11}}$; north DOP or $NDOP = \sqrt{H_{22}}$; vertical DOP or $VDOP = \sqrt{H_{33}}$; and time DOP or $TDOP = \sqrt{H_{44}}$. These DOPs are sometimes combined to form new DOPs such as the total geometry DOP or $GDOP = \sqrt{H_{11} + H_{22} + H_{33} + H_{44}}$; the three-dimensional position DOP or $PDOP = \sqrt{H_{11} + H_{22} + H_{33}}$; and the two-dimensional horizontal positioning DOP or $HDOP = \sqrt{H_{11} + H_{22}}$. Note that in many land applications, when the receiver can be assumed to be on the ground and its height constrained, the HDOP will improve.

DOPs can be viewed as the link between the pseudorange errors and PVT estimation errors. Since DOPs change as the user-satellite geometry changes over time, this implies that a given level of pseudorange measurement error will translate into different levels of PVT errors. The range of DOP values under different conditions and how they affect the position accuracy can be explored using the GNSS simulator, as demonstrated with a number of examples below.

A more detailed discussion is beyond the scope of this book. However, the book does discuss how DOPs are actually used in real-world applications. In particular, Chapter 10 shows how VDOP is a crucial metric for real-time integrity monitoring of precision landing systems. Included with the examples later in this chapter we will illustrate how the GDOP can change significantly over time and space. It turns out that the DOP value is a good general indicator to use when quality monitoring of the receiver outputs is required.

3.2.4 Estimating Receiver Velocity and Clock Drift

GNSS velocity estimates can be generated using the least squares approach with the derivative of (3.2). That is,

$$\dot{\rho}_c^k = \dot{r}^k + \dot{b} + \dot{\epsilon}_T^k \quad (3.15)$$

This is the starting point for deriving the velocity equation. Thus, following logic similar to that used in developing the position solution, this equation can be written in terms of satellite-to-receiver relative velocities and line-of-sight unit vectors. The resulting equation can be expressed as:

$$\dot{\rho}^k = [\mathbf{v}^k - \mathbf{v}_u] \cdot \mathbf{L}_{unit}^k + \dot{b} + \dot{\epsilon}_T^k \quad (3.16a)$$

where,

$\dot{\rho}^k$ is the pseudorange rate for satellite k (measured in meters/seconds). Note that the c subscript has been dropped, as pseudorange rate measurements may not be corrected before being used.

\mathbf{v}^k is the velocity vector for satellite k (normally calculated from navigation data, meters/seconds).

\mathbf{v}_u is the velocity vector for the receiver (unknown, meters/seconds).

\mathbf{L}_{unit}^k is the receiver-to-satellite line-of-sight unit vector from (3.6e).

\dot{b} is the rate of change of the receiver clock or clock drift (unknown, meters/seconds).

$\dot{\epsilon}_T^k$ is an accumulation of the unknown pseudorange rate error terms (meters/seconds).

This equation can be manipulated according to the steps below in order to arrive at an equation for the receiver velocity and clock drift without iteration. To do this, first group all known or measured quantities to one side as follows:

$$\dot{\rho}^k - (\mathbf{v}^k \cdot \mathbf{L}_{unit}^k) = -(\mathbf{L}_{unit}^k \cdot \mathbf{v}_u) + \dot{b} + \dot{\epsilon}_T^k \quad (3.16b)$$

Rewriting this in terms of the geometry matrix results in:

$$\dot{\rho}^k + \mathbf{G}_k \begin{bmatrix} \mathbf{v}^k \\ 0 \end{bmatrix} = \mathbf{G}_k \begin{bmatrix} \mathbf{v}_u \\ \dot{b} \end{bmatrix} + \dot{\epsilon}_T^k \quad (3.16c)$$

where \mathbf{G}_k is the k th row of the \mathbf{G} matrix. Assuming sufficient measurements are present, it is possible to form a system of equations and solve for the least squares estimate of \mathbf{v}_u and b_k from the following:

$$\begin{bmatrix} \mathbf{v}_u \\ \dot{b} \end{bmatrix} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{T} \quad (3.17)$$

where

$$\mathbf{T} = \dot{\mathbf{p}}^k + \mathbf{G}_k \begin{bmatrix} \mathbf{v}^k \\ 0 \end{bmatrix}$$

Using (3.17), solving for the receiver velocity and clock drift is straightforward, as will be demonstrated later using simulated data input into the MATLAB/Octave and Python scripts that come with this book. Figure 3.3 illustrates the steps involved in estimating a receiver's velocity and clock drift.

The steps in the process can be summarized as follows.

- Calculate the \mathbf{G} matrix in the exact same manner as was done for the position solution. As a matter of fact, in actual implementation of the point solution

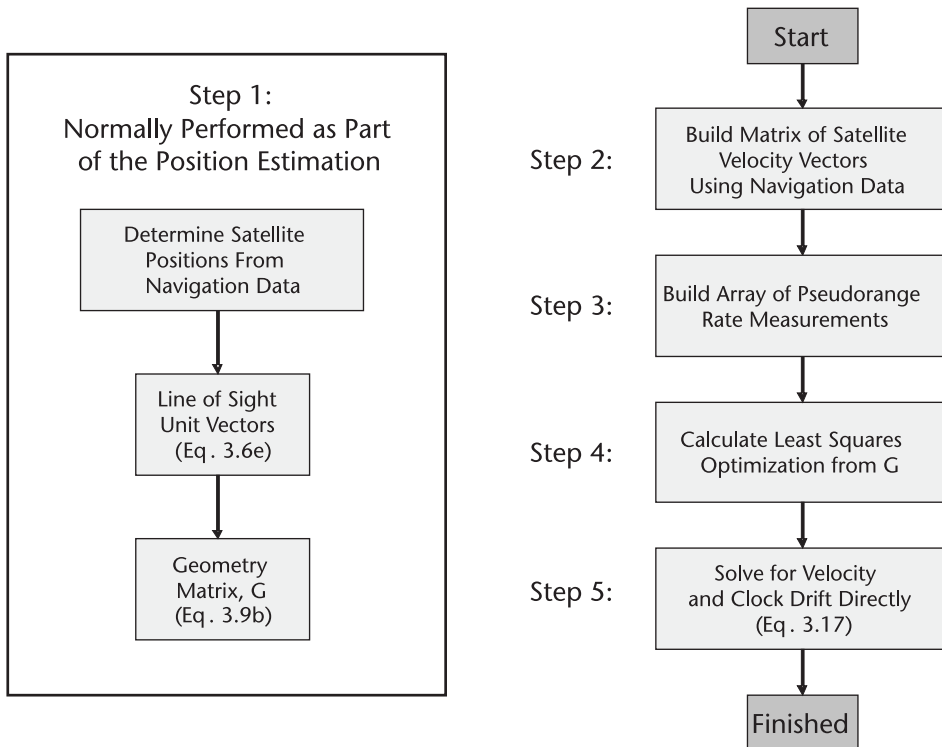


Figure 3.3 Illustration of calculation steps in the velocity calculation.

approach, it may have already been computed as the receiver velocity is normally calculated directly after the position.

- Build the satellite velocity matrix \mathbf{V}_s using the satellite velocity vectors determined using the satellite navigation data or external ephemeris information.
- Convert the raw receiver Doppler frequency measurements into pseudorange rates. This is only necessary if Doppler frequency measurements are being used. It is also possible to use a direct output of a pseudorange rate measurement if provided. Converting a Doppler frequency measurement to a pseudorange rate is straightforward and is done using the following relation:

$$\dot{\rho}_c^k = \frac{cD}{f}$$

In this relation, D is the measured Doppler frequency, f is the nominal signal transmit frequency, and c is the speed of light.

Alternatively, the pseudorange rate can be computed by differentiating the carrier phase measurements over a short interval (typically 1s). When robust carrier phase tracking is achieved this will usually result in more accurate measurements than the output of the carrier tracking loop.

- Calculate the \mathbf{T} vector from individual satellite pseudorange rates, satellite velocities, and geometry.
- Estimate the receiver velocity and clock drift rate directly by substituting the above arrays/and matrices into (3.17).

3.2.5 Estimating Time

The calculation needed to estimate the time the receiver measurements were taken is already partially completed. That is, that the receiver clock bias b has already been calculated as part of the position solution. All that is needed is to determine the absolute time to which this bias value is referenced. This is done using data in the satellite's navigation data message. Within the navigation data stream is a GPS timestamp for each transmitted subframe. This information is primarily used by the receiver to make pseudorange measurements, but it can also be used to coarsely set the receiver reference clock. Therefore, the precise GPS time when the receiver latched the pseudorange and Doppler measurements can be expressed as,

$$t = t_{ref} + b \quad (3.18)$$

where t is the signal reception time [as introduced in (3.1)], t_{ref} is the initial GPS reference time decoded from the satellite's navigation message, and b is the receiver clock bias as calculated in Section 3.2.1. The reference time is expressed in the form of GPS week and GPS seconds into the week and can be converted to any common timing standard (e.g., UTC or local time) using well-known functions. Included on the DVD are Python routines (see Chapter 13, Space Applications Processing_GRACE_Data on the DVD), which converts GPS time (in the format of week and seconds) into UTC time (day, month, year, hour, minutes, and seconds). The reader interested in more details on GPS time and all the subtle issues associated

with it are encouraged to consult the very good explanations in either Chapter 3 of [1] or Chapter 2 of [2].

3.2.6 PVT Estimation Using an Extended Kalman Filter (EKF)

The estimate of the PVT can also be obtained by using an extended Kalman filter (EKF). Conceptually, the EKF-based solution is not very different from the least squares solution presented earlier. This is because the EKF is, to large extent, nothing more than a weighted, recursive least squares estimator [10]. A notable difference between a least-squares point solution discussed above and that of an EKF is that an EKF assumes some knowledge of the receiver dynamics. When these assumptions are incorrect, problems can arise quickly. However, when they are reasonable the outputs from an EKF will often be better than those from the least squares method. A MATLAB/Octave script (GNSS_Sim_PVT_EKF.m) has been included on the DVD, which will estimate position and velocity using an EKF and simulated data generated using the GNSS simulator described in Section 3.3. Chapter 6 details the EKF-based PVT estimator.

3.2.7 Enhanced Accuracy via Carrier Phase Positioning

Precise positioning using carrier phase measurements will not be discussed in detail here because a more in-depth treatment of it is provided in Chapters 7 and 14. The process of determining position using carrier phase is computationally more difficult and fraught with many pitfalls. However, the achievable positioning accuracy is on the order of millimeters to decimeters, which is a great improvement over the code phase solution. This is because GPS carrier phase can be measured with accuracy on the order of millimeters to centimeters as opposed to the decimeters to meters accuracy of a code phase measurement. This improved accuracy allows the significant improvement in the position estimation.

Briefly, the difficulty in performing a carrier phase based pseudorange measurement primarily lies in the challenge of a quick and robust determination of the carrier cycle ambiguity for each satellite tracked. This ambiguity can be thought of as the exact number of cycles contained in the space between the receiver and the satellite at the initial signal lock. In the case of the GPS L1 signal, the carrier wavelength is only 19 cm. The separation distance between terrestrial users and satellites is approximately 2×10^7 meters for a ground-based receiver. This translates to a very large number of integer carrier cycles. Fortunately, carrier phase positioning is performed in a differential mode, and, thus, the search space for integer ambiguities is smaller than that contained in 2×10^7 meters. The process of estimating and validating the correct estimate of the integers, however, can be a time-consuming process. Furthermore, if for some reason the receiver loses carrier phase lock, even for an instant, the carrier cycle ambiguity must be recalculated.

3.2.8 Error Sources

This section briefly describes significant errors present in real measurements as well as those generated by the GNSS simulator described in Section 3.3. More

thorough explanations of the various errors in GPS measurements, supported by empirical data, are included in other GPS textbooks. A particularly comprehensive discussion can be found in Chapter 4 of [1]. For this brief overview it will be sufficient to start with the pseudorange measurement equation (3.1) and discuss the errors associated with each of the different terms.

3.2.8.1 Range Term, r^k

This term includes the position of the GNSS satellites as determined from the satellite navigation message. The orbits of the GPS satellites are modeled very precisely by the GPS control segment, but there will always remain a small error, which will vary slightly over the update cycle.

3.2.8.2 Satellite Clock Term, δt^k

Also included as part of the satellite navigation message are the clock correction coefficients for the transmitting satellite. These corrections are very good but not perfect and will result in a small error on the pseudorange. More sophisticated differential techniques can be used to improve the satellite clock error estimate and, thus, mitigate the PVT inaccuracies introduced by this error.

3.2.8.3 Ionosphere Term, I^k

The ionosphere consists of the area between approximately 50 and 100 km above the Earth. A delay added to the code phase signal while it is traversing this region can vary considerably depending on the time of day and season. These delays can range up to and over 10m in extreme conditions. The model developed by Klobuchar [3] is a good first approximation for these delays.

3.2.8.4 Troposphere Term, T^k

Broadly, the troposphere is everything under the ionosphere. In reality the error causing part is almost exclusively concentrated below 20 km and consists of dry gases and water vapor. The errors added to GNSS signals as they pass through the troposphere are less than those of the ionosphere but can still significantly impact the accuracy of position estimation. The troposphere errors are a function of atmospheric temperature and pressure in the vicinity of the receiver. To some extent, this error can also be mitigated with models.

3.2.8.5 General Range Errors (Including Receiver Noise and Multipath), ϵ^k

Noise added as part of the down-conversion and tracking of the signals during measurements is inevitable. However, with modern receivers these errors can usually be kept to manageable levels.

Multipath is the reflection of GNSS signals by an external reflector located near the receiver antenna. This reflected signal gets mixed with the direct signal. It enters the receiver processing chain and distorts the correlation tracking peak,

thereby adding an error to the pseudorange measurement. These errors can be quite large in severe multipath environments (in dense urban areas for example), and there is no sure way to eliminate them.

Chapter 4 discusses error sources in more detail, including a summary in Table 4.1 of the expected magnitude of all the terms listed above. For those interested, a good discussion of the GPS error sources and positioning accuracy can be found in the *GPS World* article “GPS Accuracy: Lies, Damn Lies and Statistics” [4].

3.3 GNSS Simulator

Included on the DVD accompanying this book is a GNSS constellation and receiver simulator. This simulator and the associated processing scripts were designed using freely available tools and libraries and will all run on Linux, Macintosh, and Windows. The details of its design, configuration, and operation are described in this section. The source code is included and can be reused under the terms of the GNU general public license, a copy of which is included on the DVD.

3.3.1 GNSS Simulator Measurement Details

With some of the basic theory behind GNSS positioning covered, it is now possible to walk the reader through some hands-on examples. This is done using a tool developed to generate practical (and possibly entertaining) examples. The tool is a GNSS simulator, which can be run to generate examples of the PVT solution. The simulator will generate pseudorange and Doppler measurements that can be used to illustrate the various steps in the calculations by postprocessing log files using a MATLAB/Octave or Python script.

3.3.1.1 Simulated Pseudorange Measurements

The simulator’s measurement generation process can be described by going through (3.1) term-by-term and discussing the implementation of the pseudorange calculations. How each term in (3.1) is handled by the GNSS simulator is described briefly below.

- $r^k(t, t - \tau)$: This is the true or geometric range and is calculated as the magnitude of the vector between the simulated receiver position at time t and the simulated satellite positions at time $t - \tau$. The satellite position is adjusted to account for the Earth’s rotation during the “transmission” time as in (3.9).
- $c\delta t_u(t)$: This is the receiver clock offset. It is a time-varying quantity computed from:

$$c\delta t_u(t + 1) = c\delta t_u(t) + c\dot{b}_s + c\dot{w}_s \times wgn(0, 1)$$

It can be seen that it is modeled using a simulated “true” receiver clock drift \dot{b}_s and a random walk magnitude term \dot{w}_s , integrated over one-second

timesteps [where $wgn(0, 1)$ is white Gaussian zero mean noise with a standard deviation 1].

- $c\delta t^k(t - \tau)$: This is the satellite clock offset relative to GPS time at the time of transmission. It is set to zero by default but the simulator provides the option of adding a constant offset.
- $I_{L1}^k(t)$: This is the delay due to the ionosphere. The default setting is to apply the Klobuchar delay model [3], based on the satellite elevation with respect to the receiver and the local time of day. Alternatively, the user can specify the zenith delay as a constant to which the obliquity factor is applied. Another option available is to set to a constant the individual values of delay can be set for each satellite.
- $T_{L1}^k(t)$: This is the delay due to the troposphere. This error is normally divided into dry and wet delays, which can be scaled with respect to the satellite elevation angle. By default the zenith dry delay is calculated as a function of receiver latitude and altitude using the Saastamoinen model, while the zenith wet delay (usually significantly less than the dry delay) is set to a constant value of 10 cm. The Saastamoinen function for the zenith wet delay can be found in [1], while the total pressure is calculated as a function of height using the standard mean sea level pressure [5]. Additionally, as in the case of the ionospheric delay, the wet and dry tropospheric delays can be specified as constants multiplied by an elevation function or individually applied to each satellite.
- $\epsilon^k(t)$: This is random noise on the pseudorange measurements. In the simulator it is modeled by the following relation:

$$\epsilon^k(t) = \epsilon_{\eta}^k \times wgn(0, 1) + \epsilon_{\beta}^k$$

It can be seen that it includes a Gaussian random noise of magnitude term ϵ_{η}^k and a constant bias value ϵ_{β}^k (which by default is zero) for each satellite in the configuration file. The bias term could be used to add a multipath error, for example. More detailed information on how to configure the simulated atmospheric errors can be found on the GNSS_Sim_Details.txt file described below and included on the DVD.

In order to demonstrate the benefit of dual-frequency measurements, the option to output a second pseudorange at a different frequency has been included in the simulator. The purpose of this second measurement is solely to demonstrate the possibility of reducing the ionospheric error present on the first pseudorange. For this reason it is calculated very simply as a function of the first pseudorange ρ_{L1}^k , the ionospheric error on this measurement I_{L1}^k , and the two signal frequencies. In the case of this particular GNSS simulator, these frequency values are given default values of $f_1 = 1575.42$ MHz = GPS L1 and $f_2 = 1176.45$ MHz = GPS L5. The resulting equation for the second frequency pseudorange measurement is,

$$\rho_{f_2}^k = \frac{I_{L1}^k(f_1^2 - f_2^2)}{f_2^2} + \rho_{L1}^k \quad (3.19)$$

Note that it is not accurate to include the same measurement noise in the two signals ϵ_{η}^k ; therefore, a recalculation of the noise for the second pseudorange is necessary.

3.3.1.2 Simulated Doppler Measurements

The calculation of the direct Doppler frequency is straightforward, the basics of which can be found in any basic physics textbook. In the case of the simulator, the method for calculating the simulated signal Doppler frequencies involves a straightforward calculation based on the receiver and satellite-simulated position and velocity vectors as well as an approximation of the rate of the receiver clock drift. The expression used for this calculation is:

$$D^{direct} = (\mathbf{v}_u - \mathbf{v}^k)^T \frac{(\mathbf{x}^k - \mathbf{x}_u)}{\|\mathbf{x}^k - \mathbf{x}_u\|} \frac{f_1}{c} + D^{clk} + \epsilon_D^k \quad (3.20)$$

where:

\mathbf{v}_u is the receiver velocity vector.

\mathbf{v}^k is the GNSS satellite velocity vector.

\mathbf{x}_u is the receiver position vector.

\mathbf{x}^k is the GNSS satellite position vector.

f_1 is the primary simulated GPS frequency (normally GPS L1).

c is the speed of light in a vacuum.

$D^{clk} = \left(\frac{\dot{b}}{c}\right) f_1$ is the Doppler contribution due to the GPS receiver clock drift \dot{b} .

$\epsilon_D^k = \epsilon_{D\eta}^k \times wgn(0, 1) + \epsilon_{D\beta}^k$ includes the specified error terms to be applied to the simulated L1 signal Doppler frequency measurement. That is, it includes a Gaussian random noise magnitude term $\epsilon_{D\eta}^k$ and a constant bias value $\epsilon_{D\beta}^k$ for each satellite in the configuration file.

Within the equation above, the clock drift of the GPS receiver can be significant and will add an unwanted error to the ideal Doppler calculation. However, the value of this drift is estimated during the receiver velocity estimation detailed earlier.

3.3.2 GNSS Simulator Interface Files

3.3.2.1 Input Files

The GNSS simulator can be configured in several ways depending on the reader's desire and inclination to explore deeper into the details. However in all cases the user will interface with it via the following input and output files.

- A GNSS satellite initialization file (GNSS_Sim_Default_TLE.txt). A default file is loaded automatically corresponding to 12:00:00 Aug 17, 2007. This file includes the two-line orbital elements as provided by NORAD [6] for all the operating GNSS satellites at this time. These elements are used to propagate the locations of the GNSS satellites around a model Earth upon which it is possible to place GNSS receivers. Updates for the NORAD two-line elements (TLEs) can be obtained from the Internet CelesTrak Web site [7] using the interactive Python script included (GNSS_Sim_GetTLE.py). Additionally, a modified version of the TLE file can be loaded that contains preliminary entries for the Galileo constellation (which can be generated using the GNSS Simulator, “Features→Create TLEs”). This is useful to demonstrate the potential improvement in coverage provided by increasing the number of visible satellites (and assuming that the receiver is capable of processing all the visible signals). This file is can be modified by the user as desired. For example it can be modified to add GLONASS satellites in addition to GPS or Galileo satellites.
- Receiver operating scenario information (GNSS_Sim_Scenario1.txt). There are two options for simulating user GNSS receivers, a terrestrial scenario and a receiver located on board a Low Earth Orbit (LEO) satellite. For the terrestrial receiver only, the position and velocity can be controlled as a function of time by loading a scenario file. If the initial position and movement is not specified in the loaded scenario file then the conditions specified on the main page of the GNSS simulator are used. The scenario capability exists to enable automated programming of journeys around the simulated Earth environment. There are three types of possible receiver movement scenarios: (1) simple position and velocity commanding, (2) destination journey planner, and (3) Waypoint journey planner. Each of these methods is illustrated with an example in the following sections. (See the example text files for detailed information).
- Simulation details (GNSS_Sim_Details.txt). The default file containing these values is automatically loaded and requires no specific action from the operator. The parameters in this file include typical values for the internally simulated clock parameters on the receiver (cb_s and $c\dot{w}_s$) and the GNSS satellites as well as reasonable noise and bias values to place on the pseudo-range (ϵ_{η}^k and ϵ_{β}^k) and Doppler ($\epsilon_{D\eta}^k$ and $\epsilon_{D\beta}^k$) measurements that are output by the simulator for postprocessing. Additionally, the atmospheric errors applied can be specified either as constant biases or modeled as described above. The exact entry formats for this file are detailed in the header information of the file listed above. An interesting exercise that can be performed using this file is to make the measurements, either perfect (setting all errors to zero) or unreasonably bad (entering very large biases) to illustrate the range of results that can be obtained by varying the different measurement parameters.

3.3.2.2 Output Files

Simulator output files are generated separately for the terrestrial and LEO receivers depending on the logging configuration. For the LEO receiver logging is set to

“OFF” by default. The data output by the GNSS simulator can be configured using the “Features→Configure Logging” menu. It will usually be the case that only a single receiver is of interest to a particular user (ground based or LEO) and it would not be useful to always generate two output files. The standard output files are listed as follows:

1. Raw measurement output file, near-Earth receiver (GNSS_Sim_Output1.dat);
2. Raw measurement output file, LEO spacecraft receiver (GNSS_Sim_Output2.dat, default OFF);
3. GoogleEarth position output file for near-Earth receiver (GNSS_Sim_GoogleEarth1.kml);
4. GoogleEarth latitude and longitude output file for the space-based receiver (GNSS_Sim_GoogleEarth2.kml; default OFF).

Every row of the output files described by (1) and (2) above corresponds to a timestep. At every timestep, all the reference and measurement values necessary to repeat the PVT calculations described above are logged. This includes raw pseudorange and Doppler measurements, GNSS satellite positions and velocities, and simulated position estimates and truth conditions for use in comparison. A detailed column breakdown describing of the output files is included at the start of the postprocessing MATLAB/Octave and Python scripts.

Output files (3) and (4) are interfaces to Google Earth. They are Google Earth loadable kml files generated with the altitude set to 0 for the space based receiver. These files can be loaded into Google Earth (after starting Google Earth separately, “File→Open”) and will display a series of place markers at every output (corresponding to the step size) over the duration of the simulation.

3.3.3 Postprocessing GNSS Simulator Output Files

Provided on the DVD are three postprocessing scripts: one written in MATLAB/Octave (GNSS_Sim_PVT.m), a second in Python (GNSS_Sim_PVT.py), and an additional MATLAB/Octave script that uses an EKF solution to estimate the position and velocity of a simulated receiver. MATLAB is a widely used high-level computing language that excels at numerical computation and plotting. However, MATLAB may be prohibitively expensive for many readers and throughout this book efforts have been made to ensure that all scripts were compatible with the open source program Octave as well.

To postprocess the GNSS simulator output it is often easiest to copy the output files listed above into the script directory. When the scripts are executed, the user will be prompted as to which data set to process (near-Earth or satellite-based receiver). Next, if the terrestrial receiver was selected, users will be asked if they would like to recalculate the position solutions using the raw measurements with single- or dual-frequency measurements. Afterward, a series of plots are produced and all the intermediate variables are made available in the workspace for inspection.

The examples below all utilize these postprocessing script files to display the results. The reader is encouraged to make local copies of the scripts and inspect and plot intermediate variables during the calculations (for example, try calculating the HDOP and VDOP separately and comparing them with the horizontal and vertical positioning results).

3.4 GNSS Simulator Examples

As noted in the introduction, a motivation for writing this book was to give readers hands-on experience with GNSS algorithms. Consistent with this philosophy, in this section we present a series of examples, which we believe will underpin the brief theoretical sketch given in the first half of this chapter or the more detailed treatment found in [1].

3.4.1 Example 1: Simple Navigation

The initial example will involve a single ground based receiver located in Guildford, United Kingdom. To make the example a little bit more interesting, a scenario has been created that moved the receiver in an approximate square while hovering at 100m (ignoring roads and local obstacles) around the starting point. To run this example, start the GNSS simulator and hit the “start” button. You are now ready to control your receiver and output raw measurements.

By hitting “start” the satellite TLE entries (default entries in `GNSS_Sim_Default_TLE.txt`) and the `GNSS_Sim_Details.txt` files will be automatically loaded and used to configure the GNSS satellites and simulation time (based on the satellite receiver TLE reference time). Figure 3.4 shows the default receiver location in the main display. Figure 3.5 shows an example of the detailed maps available, in this case with the terrestrial receiver moved to North America, where this scenario could also be run with similar results. These receiver locations were influenced by the lead author’s personal bias, but there are no constraints as to where on Earth the user receiver could be placed. The location can be changed manually using the display or with an entry in a scenario file. The reader is encouraged to experiment with different user locations.

It is now possible to step through time (using the “step” button and interval located below and to the right of the main map) and simultaneously generate an example PVT solution (displayed in the “Features→Estimated Position” menu), while also outputting raw measurements that can be used to repeat the PVT calculation afterward. As mentioned, in addition to manually controlling the receivers over time, it is possible to load a movement scenario for the near-Earth receiver. This is done using the “load scenario” button, and selecting the `/Simulations/Example1/GNSS_Sim_Scenario_Example1.txt` file. At this point the scenario is active and all manual velocity adjustments using the display will be ignored. Now, as you step over time, the receiver will move as specified in the scenario file, with its velocity changing according to the entries in the file (in this case every 60 seconds, turning left 90 degrees). Alternatively, the whole scenario can be accelerated quickly over 300 seconds using the “go” button under the “run over interval” heading.

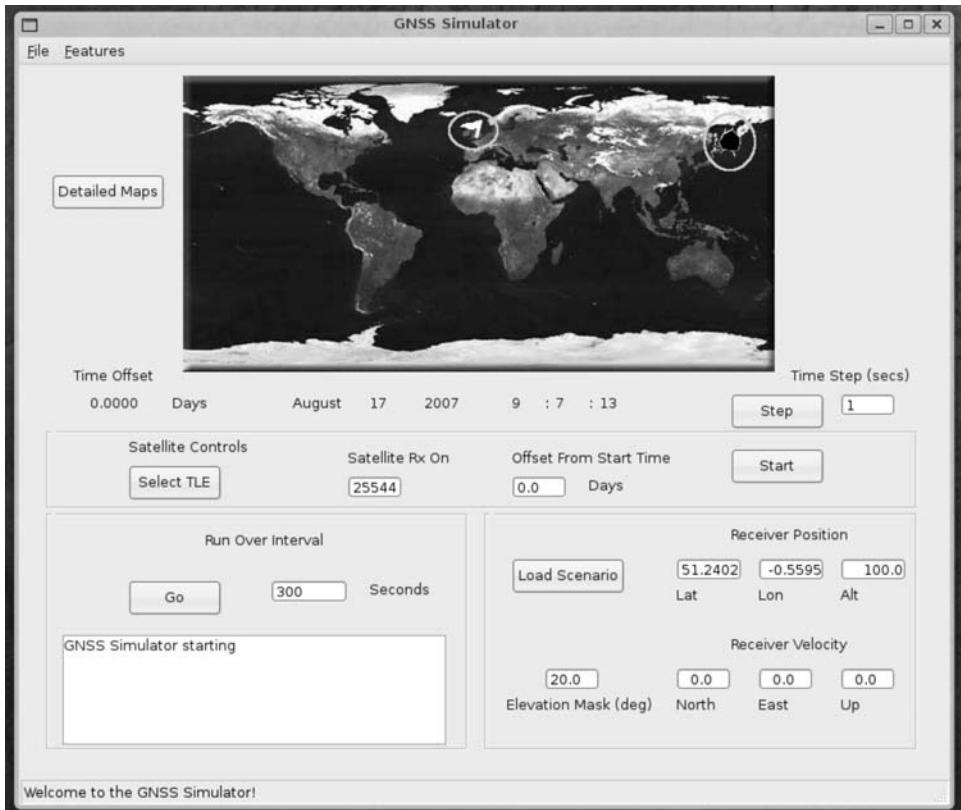


Figure 3.4 Main display of the GNSS simulator contained on the CD accompanying this book. The terrestrial receiver is represented on the world map as a flag.

This will run the simulation over the time interval specified (using the timestep specified in the scenario file). After the interval has been completed the near-Earth receiver will have moved according to the entries in the scenario file. At this point the simulator reverts to manual control, where the user can continue to step through time or exit and postprocess the output file.

In order to start postprocessing the data, run the MATLAB/Octave or Python script in the same directory as the output file (GNSS_Sim_Output1.dat), after adding it to the search path. When the script is run it will produce several plots, including the two shown in Figures 3.6 and 3.7 (which has been divided into three parts to be clearly visible in black and white) of the receiver position and velocity over the duration of the simulation.

As mentioned above, the position estimates as determined within the GNSS simulator can also be plotted as place markers in Google Earth. For this example, the receiver moving as commanded around the approximate shape of a box, in the vicinity of Stoke Park in Guildford, United Kingdom is shown in Figure 3.8.

3.4.2 Example 2: Traveling Between Destinations

The next example is a simulated journey between London and Edinburgh in the United Kingdom. Larger scenarios such as this can be useful for looking at the



Figure 3.5 Detailed display from the GNSS simulator showing another example of a simulated terrestrial receiver near Buffalo, New York, and a satellite receiver off the west coast of Mexico. Most of the GPS satellites used in the receiver simulation are outside the map frame; however the lat/lon subsatellite points of several GPS satellites are shown.

estimated satellite coverage over longer time intervals and greater distances. However, more attention needs to be paid to the step size and receiver velocity. In order to keep the receiver output files from growing to unmanageable sizes, the simulation step needs to be better controlled. Thus, it is not desirable to log every second of a trip from London to Edinburgh when every 5 minutes or even longer will suffice. Additionally, to move the receiver from place to place at a given speed, the time between changes in direction needs to be properly calculated to ensure that the receiver is able to reach the desired destination in the desired time. The destination and velocity scenario option has been designed for these cases. This option provides for time-based scenario entries that specify the receiver destination coordinates (with both horizontal and vertical velocity components) and a simulation time step.

The GNSS simulator is started in exactly the same manner as in Example 1. However, this time, load the scenario file `/Simulations/Example2/GNSS_Sim_Scenario_Example2.txt`. The scenario consists of three steps: (1) navigating the receiver from its initial location (Guildford) to central London, and then (2) from London to central Edinburgh in large steps, and finally (3) slowing down and shortening the timestep before final arrival near the Edinburgh town center.

To run the entire scenario, load the scenario file as described above and specify the durations as 15,000 seconds. Then press the “go” button under the “run over interval” text to simulate the entire journey. Note that the timestep changes

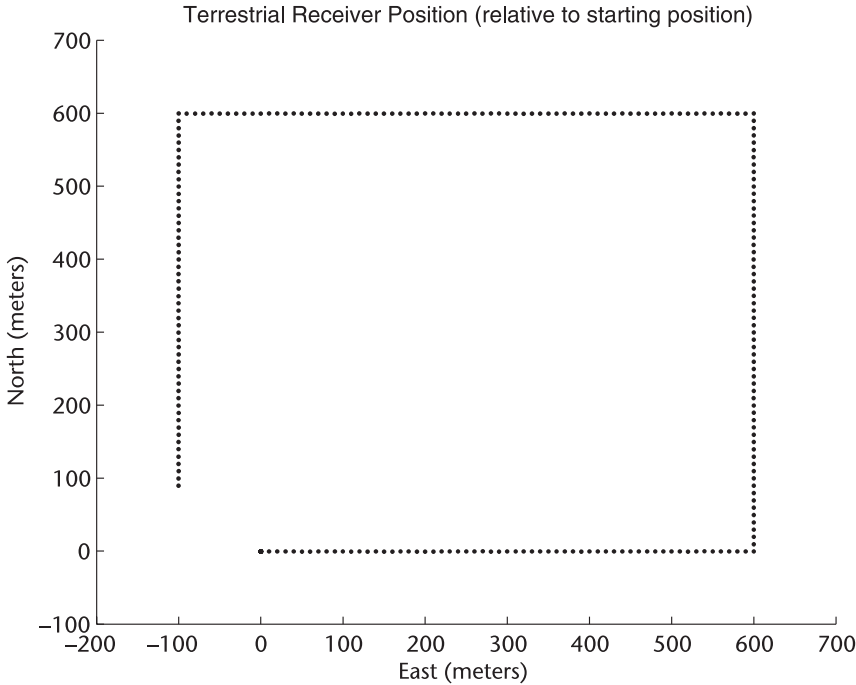


Figure 3.6 Recalculated positions of the simulated GNSS receiver over the five minutes of scenario example 1, converted into a local east-north reference frame with respect to the starting position.

automatically as specified in the scenario script. After a few seconds the scenario will complete and the receiver will automatically relocate to Edinburgh. As before, the output files can be postprocessed to reveal the details within the PVT estimation at each simulated time epoch. The resulting processing will show that, unlike the first example, the changing geometry and satellites tracked over this journey are much more varied. The GDOP estimation and satellites tracked over the duration of the scenario are shown in Figure 3.9 along with the Google Earth place markers over the simulated receiver path in Figure 3.10.

3.4.3 Example 3: Waypoint Navigation Using FlightGear

The next example uses the Waypoint scenario option to move the user linearly from point A to point B within a specified period of time. This option indicates to the GNSS simulator to calculate the trajectory and velocity needed to achieve the specified movement. This scenario option is primarily designed to use several entries at closely spaced intervals (the only requirement being that the simulated step size is smaller than, and ideally an exact multiple of, the Waypoint's command times).

To add an interesting twist to this example, the outputs of the FlightGear Flight Simulator [8] are used to demonstrate GNSS calculations during a simulated aircraft flight. In order to do this it is necessary to output the time interval and aircraft locations over the duration of a FlightGear run. The FlightGear output file used

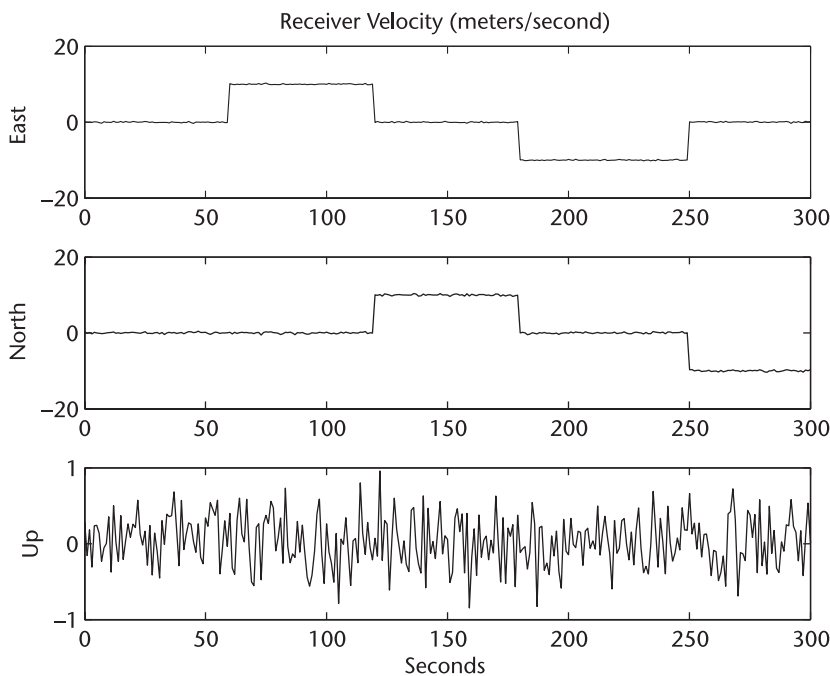


Figure 3.7 Recalculated velocities of the simulated GNSS receiver over the five minutes of scenario example 1, converted into a local east-north reference frame with respect to the starting position.

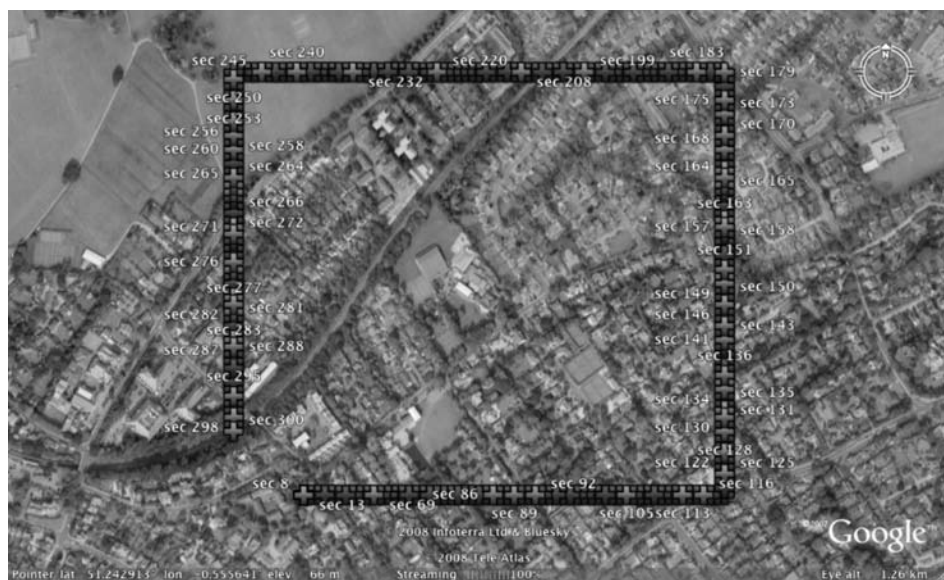


Figure 3.8 GNSS simulator estimated positions loaded into Google Earth over the five minutes of scenario example 1.

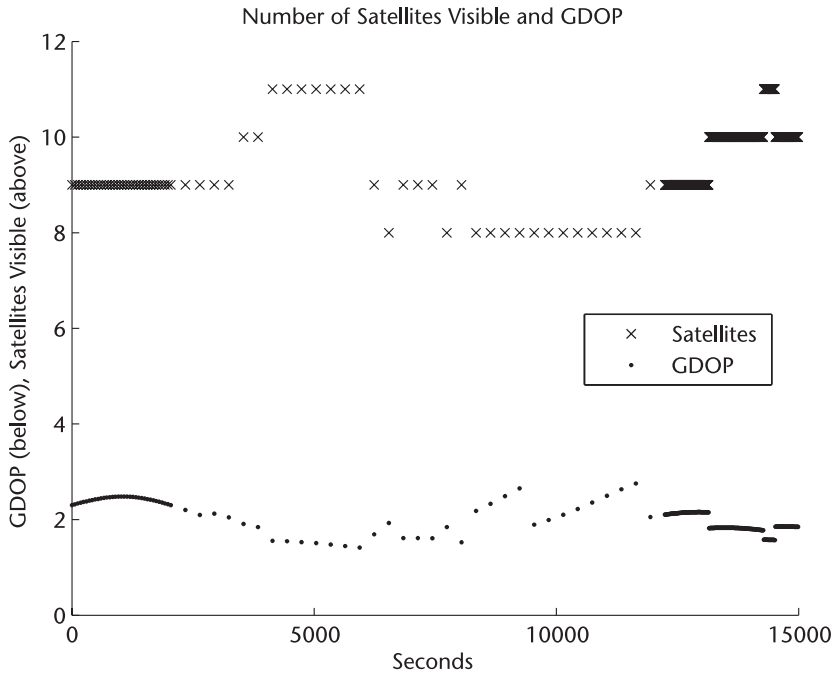


Figure 3.9 Recalculated GDOP and number of satellites visible over the entire Guilford to London to Edinburgh scenario duration of example 2.

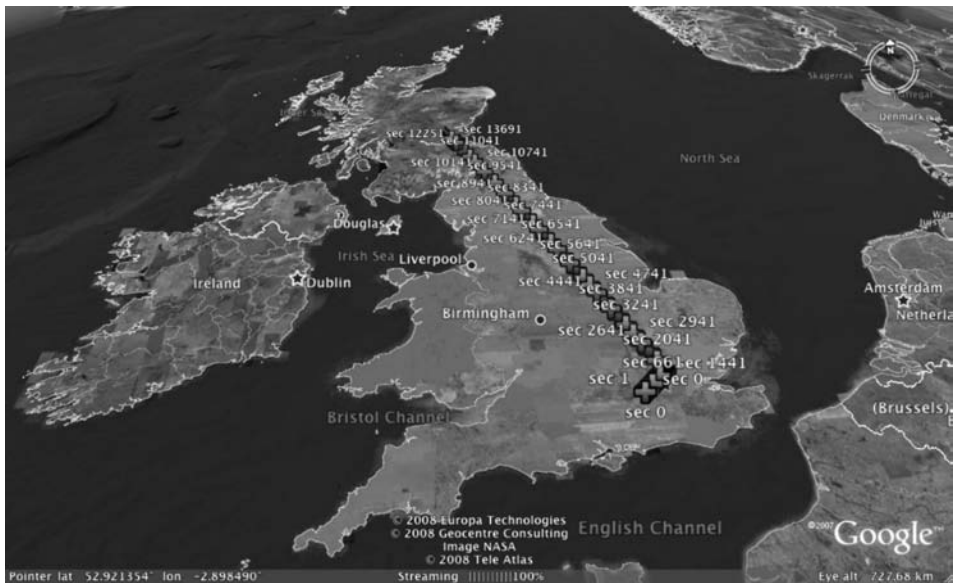


Figure 3.10 GNSS simulator estimated positions loaded into Google Earth over the entire Guildford to London to Edinburgh scenario duration.

for this example was collected during a series of maneuvers made near Chicago. The format of this file was then converted into a list of GNSS simulator Waypoint scenario entries using the “Features→Convert FlightGear Log” option in the simulator. The FlightGear output file must be comma-delimited, with the first four entries being elapsed time (seconds), longitude (decimal degrees), latitude (decimal degrees) and height (meters). If the elapsed time is not an integer second it will be rounded to keep it in sync with the simulated time step. The first entry is skipped to accommodate the optional inclusion of a one-line header, and all entries after [height] are ignored. See FlightGear_Output_Chicago.txt and GNSS_Sim_Scenario_Example3.txt for examples of corresponding FlightGear and GNSS Simulator scenario Waypoint input files (both in the /Simulations/Example3 directory).

It is then possible to run the resulting Waypoint file as a scenario as in example 2 (over a duration of 1,400 seconds). This will cause the GNSS simulator to output raw measurements as could have been expected during the FlightGear aircraft simulation. The GNSS simulator estimated flight path for this example, is shown in Figure 3.11 using the Google Earth output file.

3.4.4 Example 4: Dual-Frequency Calculation

The GNSS simulator will generate a simulated second-frequency pseudorange measurement that can be used to correct the ionospheric error present on the primary frequency measurement. This has been demonstrated by simulating a stationary receiver over the duration of an entire day. The receiver being stationary makes it possible to clearly observe the potential improvement in the positioning accuracy (as compared to the simulated truth) using the dual-frequency technique. It also



Figure 3.11 GNSS simulator–estimated positions loaded into Google Earth over the entire FlightGear flight simulator run of example 3.

allows us to observe the time-of-day dependence of the ionosphere errors as applied by using the Klobuchar model.

Running this example is quite simple because all it requires is the default setup; no scenario is needed as the receiver does not move. First, start the GNSS simulator and click on “start” to complete the initialization. Next, use the “go” button to run over an interval of one day (or 86,400 seconds), and at a 60-second timestep to keep the file sizes reasonable. The data logged to the output file will by default include L1 and L5 frequency pseudoranges. To see the difference between positioning using only the L1 frequency and the potential improvements from dual-frequency measurements it is necessary to select option “2 - Dual Frequency” when prompted during the execution of the MATLAB/Octave or Python postprocessing script. The results for both the single- and dual-frequency cases using the receiver in its default location for a single day are shown in Figures 3.12 and 3.13.

Using dual-frequency pseudoranges mitigates the error due to the ionosphere but adds noise due to the mixing of two independent noisy measurements. The mean 3D position error for the single-frequency measurements shown above is 7.26m, which improves to 4.14m, using dual-frequency methods. The remaining 4.14m of error is largely due to the simulated troposphere error.

It is necessary to stress that the method used to simulate the second frequency pseudorange is very simplistic and by no means a high-fidelity representation of reality. The primary reason for including this feature is to demonstrate how to perform the basic dual-frequency calculations and show that the positioning accuracy will improve when both pseudoranges are used. This simulator feature was intended to be used only as a coarse demonstration of the method.

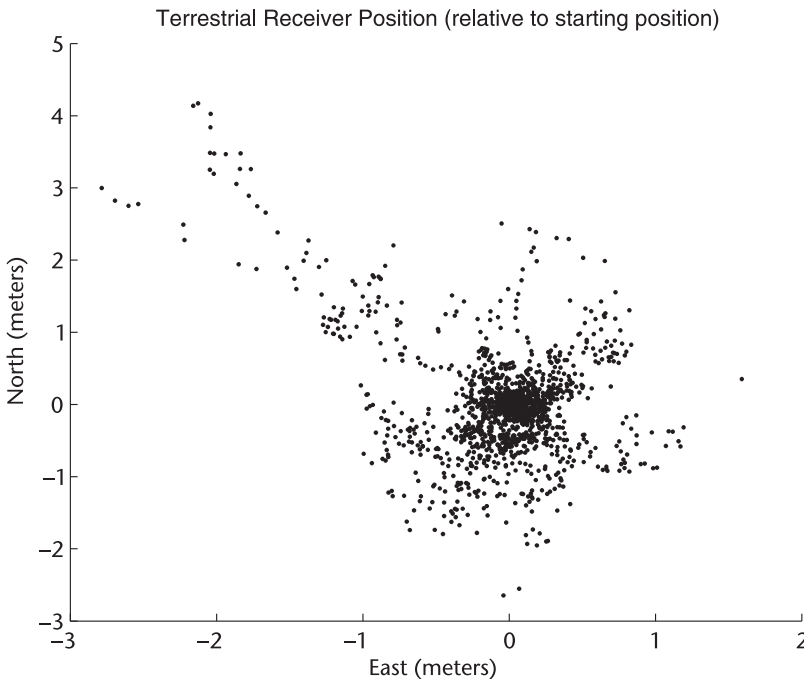


Figure 3.12 Recalculated position scatter plot over 24 hours for a stationary receiver. Position estimated using simulated single frequency L1 measurements only.

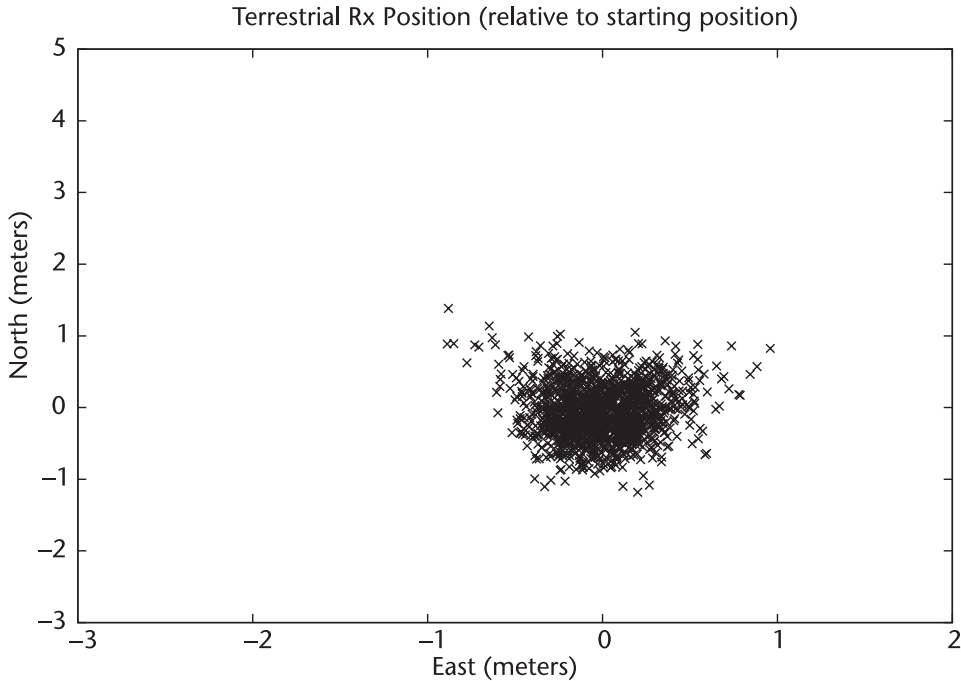


Figure 3.13 Recalculated position scatter plot over 24 hours for a stationary receiver. Position estimated using simulated dual-frequency L1 and L5 measurements.

3.4.5 Example 5: Adding Galileo Satellites

Another useful feature of the GNSS simulator is the ability to incorporate additional satellites into the simulation environment. For example, this allows us to grow the Galileo constellation from two (GIOVE-A/B) (at the time of writing) into the fully operational constellation with just a couple mouse clicks. This is accomplished by generating temporary TLE file entries, based on knowledge of the planned satellite orbits [9]. To generate TLE entries for additional satellites, open up the “Features→Create TLEs” window. This interface permits the generation of individually satellite entries or the entire Galileo constellation with a single button press (at the specified time epoch). Several entries can be created in one session, and upon exiting the GNSS simulator the created TLE entries will be written to the file named `GNSS_Sim_CreatedTLEs.txt`. From this file the entries can be cut and pasted (using any text editor) into the `GNSS_Sim_Default_TLE.txt` file or a separate file that can be loaded into the simulation. Please note that where the entry is copied into the file is important. This is because there are two types of satellite entries in the file, one set for GNSS satellites and another group for satellites to place the space-based receiver on. The dividing point is clearly marked with the text “`QQQ`”. All satellites above this line are GNSS satellites. All satellites below this text are available to place the space-based receiver on. If the reader is using his or her own TLE text file, the “`QQQ`” entry is required only if he or she wants to simulate an LEO receiver.

A file containing a full set of the planned or future Galileo satellites called `GNSS_Sim_wGalileo_TLE.txt` is provided to demonstrate the improved satellite

coverage possible when Galileo becomes operational. To use this file load it separately using the “select TLE file” button on the main display. During the manual operation of the simulation, the Galileo subsatellite locations will be displayed on the detailed Earth views as an image of the famous Italian himself.

When the scenario from example 2 is rerun, we can observe the increase in the number of satellites visible on a simulated trip between London and Edinburgh. Shown in Figure 3.14 is a plot of the GDOP and number of satellites tracked using the combined GPS and Galileo TLE file. As expected, the number of satellites tracked increases and the GDOPs across the scenario improves significantly.

The purpose of this simulation is to demonstrate the potential satellite visibility improvement by using two navigation constellations working together (the same exercise could be performed using other constellations as well, such as GLONASS).

An important consideration must be stressed in playing this game, for it assumes that the receiver being simulated is capable of making both GPS and Galileo measurements simultaneously. The difference between the two system times with respect to each other must be well determined or solved for in order to link the pseudoranges from independent systems and use them together. In reality an additional error component will be added when attempting to use both GPS and Galileo measurements together in a real application. This is dependent on how well the difference between the time and coordinate reference frames of the two constellations can be determined (Kaplan and Hegarty discuss this issue in more detail [2]).

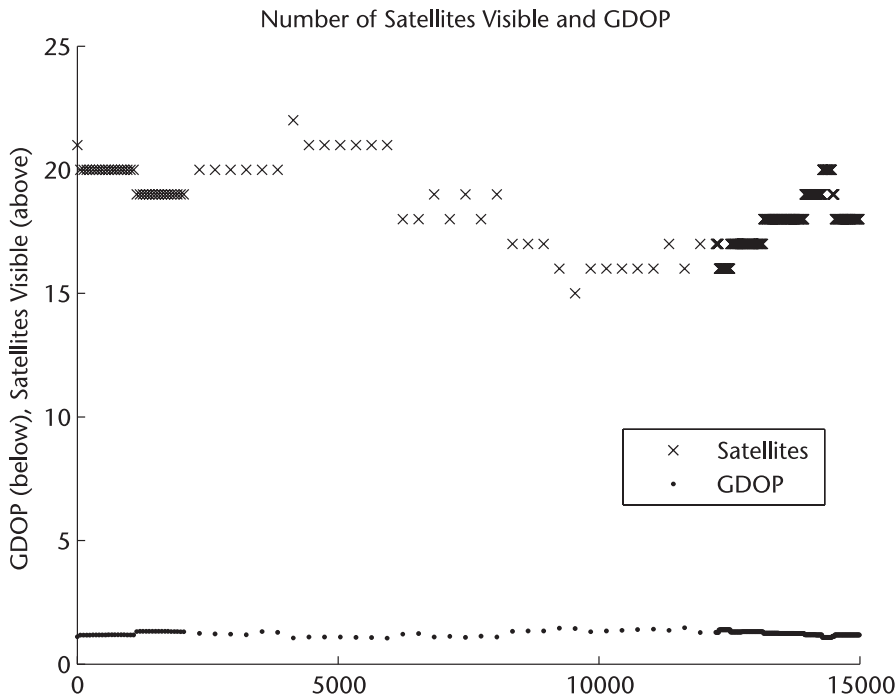


Figure 3.14 Recalculated GDOP for a simulated receiver traveling between London and Edinburgh with both full GPS and Galileo constellations visible.

3.4.6 Example 6: Spacecraft-Based Receiver

The last example will involve placing a receiver onboard a satellite in LEO. This can be done in parallel or independent of simulating a near-Earth receiver. The GNSS simulator by default logs data for only the near-Earth receiver. However, it does continually simulate a space-based receiver at the same time and initializes the simulation start time based on the satellite receiver TLE. In order to activate the logging option for the space-based receiver, use the form found under the “Features→Configure Logging” menu and set the “Log Space Based Receiver Data” option. By default, the space-based receiver is located on the International Space Station (ISS), but this can be changed to several satellites in the TLE file being used. Currently, there are a number of satellites available in the `GNSS_Sim_Default_TLE.txt` file for selection, including JASON, QuickSCAT, and SPOT 5. Others can be added as desired, both real (copied from the Web) or created using “Features→Create TLEs” (such as TestSat in the default file). New TLE entries of this sort must be added after the “QQQ” text in the TLE file to be designated as a possible GNSS receiver carrying satellite. Any added satellite entries will not appear in the drop-down list but can be selected by typing the TLE satellite ID into the “satellite receiver on” edit box directly.

When the satellite is selected and the logging turned on, the output for the satellite receiver is logged in `GNSS_Sim_Output2.dat` and can be processed exactly as the ground-based receiver. However, the rapid velocity of a satellite in LEO (typically over 7 km/s) makes for drastically different PVT results. For those interested, Chapter 13 includes a script that estimates the position of the GRACE satellite using real data. This example can now be run over a couple of satellite orbits (typically about 90 minutes in LEO) and the results plotted using the postprocessing scripts. A good combination is to set the step size to 10 seconds and use the “go” button to log results over a couple of orbits (10,000 seconds, which is approximately two orbits). When postprocessing the output files, to select the satellite receiver output file select “2 – Space Receiver” option when prompted using the postprocessing MATLAB/Octave or Python scripts. Figures 3.15 and 3.16 show the position and velocity results for this example.

3.5 Summary

This chapter presents a method for estimating PVT of a GNSS receiver. These methods are well-known and are included here mostly to provide a foundation for the GNSS simulator examples and the applications chapters presented later in this book. However, in addition to the standard derivations presented here, scripts have been developed to demonstrate how these formulas are implemented using simulated data. Our goal is to shift the emphasis in learning these methods into a hands-on context: to get the reader to put down the book, boot up their computers, and actually start estimating positions and velocities. Even though these examples all use simulated measurements, the method can be (and has been) applied exactly as demonstrated here, using real measurements (see Chapters 5 and 13 for examples).

The GNSS simulator was designed to demonstrate several aspects of existing and future GNSS receivers including dual-frequency measurements, multiple

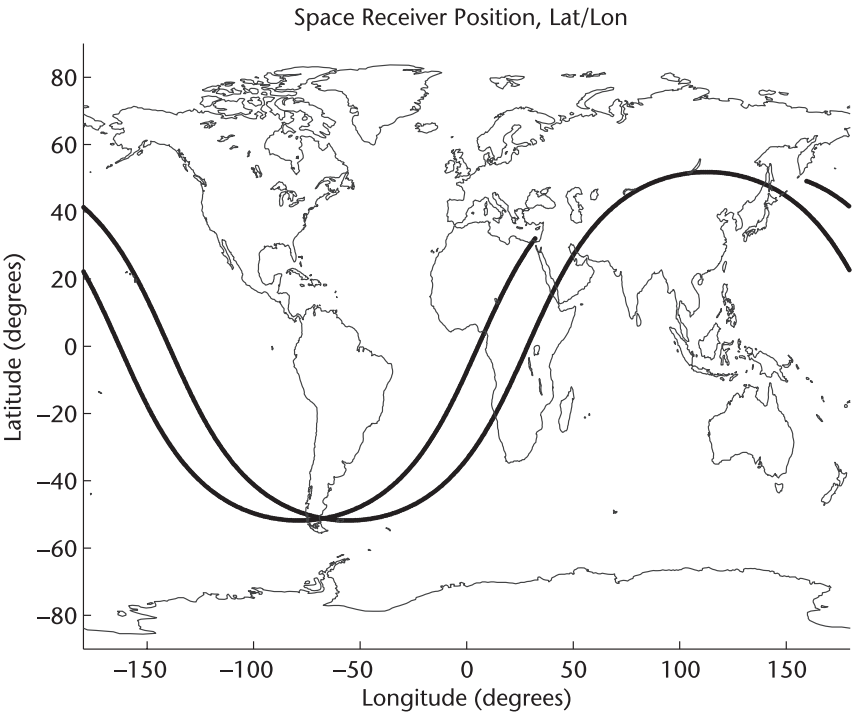


Figure 3.15 Postprocessed position for approximately two orbits from the simulated satellite-based receiver.

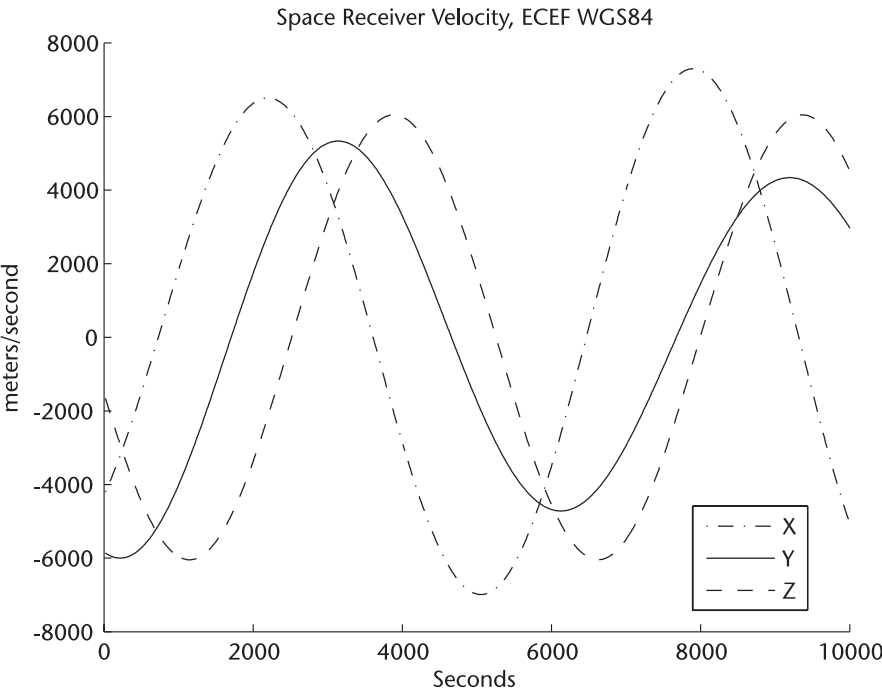


Figure 3.16 Postprocessed velocity for approximately two orbits from the simulated satellite-based receiver.

constellation measurements, near-Earth receivers in dynamic environments, and space-based receivers in LEO. Furthermore, it provides an interface with other well-known tools such as Google Earth and FlightGear.

We hope that these tools will allow readers to explore in depth the subtleties of the PVT calculations and how they vary depending on the raw measurements provided. Ultimately, we hope that readers find the utilities provided useful, and if you have a little fun (or even a passing laugh) while working at it, all the better.

3.6 Programs and Tools Provided on the DVD

The GNSS simulator and processing scripts can be found on the DVD in the subdirectory //Chapter3_Nav/. The subdirectory GNSS_Sim/Windows/ contains an NSIS installer for the Windows version of the GNSS simulator. Prebuilt versions of the simulator for Linux (built under Ubuntu 08.10), Mac OS X, and Windows XP are also included in the /GNSS_Sim directory, together with the necessary initialization files. To run the Linux version on a different distribution, download the wxWidgets libraries from the specific distribution repository and rebuild from source (using the Makefile in the /src directory). The Octave_Matlab/ and Python/ subdirectories contain the postprocessing scripts mentioned above. The Simulations/ ExampleX/ subdirectories contain the GNSS simulator output and input files for the six examples presented above.

The GNSS simulator, as well as the MATLAB/Octave and Python scripts, are all included for public reuse and distribution and released under open source software licenses (see individual files for details). Source code and Makefiles files for the GNSS simulator have been included in the GNSS_Sim/src directory.

References

- [1] Misra, P., and P. Enge, *Global Positioning System: Signals, Measurements and Performance*, Lincoln, MA: Ganga-Jamuna Press, 2001. www.gpstextbook.com.
- [2] Kaplan, E. D., and C. J. Hegarty (editors), *Understanding GPS: Principals and Applications* (Second Edition), Norwood, MA: Artech House, 2006.
- [3] Klobuchar, J. A., "Ionospheric Effects on GPS," in *Global Positioning System: Theory and Applications*, Parkinson, B., et al (eds.), American Institute of Aeronautics and Astronautics, 1996, pp. 485–515.
- [4] van Diggelen, F., "GPS Accuracy: Lies, Damn Lies and Statistics," *GPS World*, Vol. 9, No. 1, pp. 41–45.
- [5] Van Bramer, S. E., "Pressure vs. Altitude Relationships," Widener University, http://science.widener.edu/~svanbram/chem332/pdf/press_alt.pdf, 1998.
- [6] North American Aerospace Defense Command (NORAD), <http://www.norad.mil>.
- [7] Celestrak, NORAD Two Line Element Sets from the Center for Space Standards and Innovation, <http://celestrak.com/NORAD/elements/>.
- [8] FlightGear Flight Simulator, <http://www.flightgear.org>.
- [9] European Space Agency, Navigation, Galileo Space Segment, http://www.esa.int/esaNA/SEM5K8W797E_galileo_2.html.
- [10] Crassidis, J. L., and J. L. Junkins, *Optimal Estimation of Dynamic Systems*, Champion and Hall/CRC Press, 2004.

Differential GNSS: Accuracy and Integrity

Sam Pullen and Jason Rife

4.1 Introduction to DGNSS

Differential GNSS (DGNSS) was originally conceived in the 1980s as a means of removing satellite-navigation errors that are highly correlated spatially. The basic concept is to install one or more GNSS “reference receivers” at known locations and have these reference receivers broadcast estimates of GNSS satellite errors to users, who can then remove these error estimates from their own measurements. Since the largest error source on the GPS C/A-code prior to May 2000 was the deliberate clock-error distortion known as selective availability (S/A), DGNSS was first motivated by the desire to remove this artificially induced error source [1]. Now that S/A no longer exists (see [2]), DGNSS benefits users by removing almost all of the remaining spatially correlated errors that occur due to such causes as satellite clock and ephemeris error and ionospheric and tropospheric delays. DGNSS also provides a means to monitor GNSS measurements for satellite faults or other anomalies that might cause errors large enough to threaten user integrity, or safety of operations, if the measurements are not detected and removed in a timely manner.

This chapter explains the fundamentals of DGNSS, including the means by which DGNSS is implemented in practice and the expected level of accuracy resulting from DGNSS. This chapter then proceeds to describe several categories of failures and anomalies that can affect the integrity of satellite navigation and briefly explains how they can be dealt with as part of DGNSS implementation. Chapter 10 provides more details regarding the implementation of differential GPS for worldwide civil-aviation navigation.

4.2 Fundamentals of Differential GNSS

Most users of satellite navigation are so-called “stand-alone” users, meaning that they perform position determination solely from the ranging measurements made from GNSS satellites. These measurements contain errors due to satellite imperfections, delays imparted on the signals as they pass through the atmosphere, and errors due to the users’ equipment and nearby environment. The first two of these error categories changes relatively little over separations of tens to hundreds of

kilometers. Therefore, estimates of these errors, or “corrections,” broadcast by receivers at known, fixed locations can be used to almost completely cancel out satellite and atmospheric errors. This is the idea behind DGNSS. While this chapter will focus mostly on the simpler “local-area” version of DGNSS, in which scalar corrections (for all errors combined) are generated at one nearby location, it will also briefly describe the “wide-area” version of DGNSS that utilizes a widespread network of reference receivers and provides “vector” corrections for individual error sources to support users over regions of thousands of kilometers.

Figure 4.1 shows a simplified, two-dimensional diagram of a version of DGNSS intended to support users in the local vicinity (within 10–50 km) of a single reference receiver (i.e., “local-area” DGNSS). In this figure, the vectors $\mathbf{R}^{(i)}$ extend from the reference receiver to $i = 1, 2, \dots, N$ satellites in view, and the vector \mathbf{b} extends from the reference receiver to the single user shown in the picture (this vector is often called the “baseline”). The scalars $R^{(i)}$ and b represent the magnitudes of these vectors. Figure 4.1 is used to define the notation used in the analyses that follow.

In standard DGNSS implementations, the reference receiver and antenna shown in Figure 4.1 are in fixed locations. The location of the reference receiver antenna is carefully established in advance by GPS surveying techniques to within 1–2 cm (see Chapter 14) so that any errors in the estimated location of the antenna are negligible compared to the DGNSS errors described in Section 4.2.1. The use of a fixed reference station distinguishes DGNSS from another multireceiver technique known as “relative GNSS,” in which each receiver is positioned relative to a preselected “master” receiver, even though some or all of these receivers may be

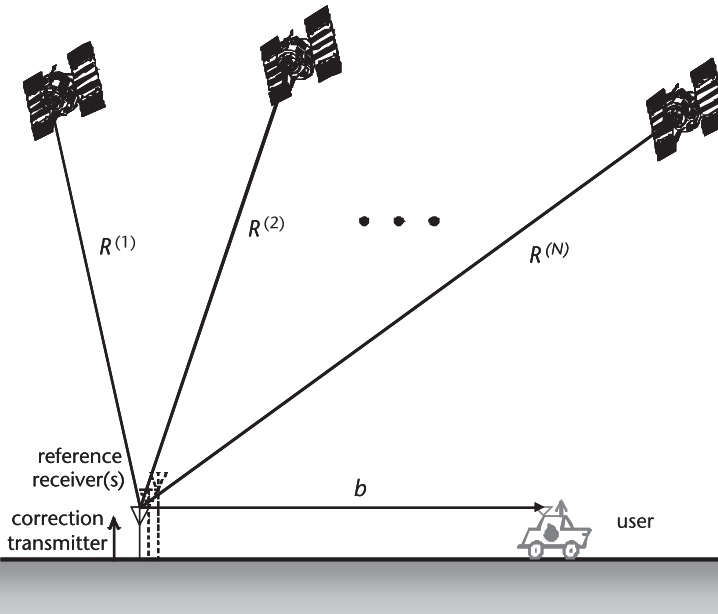


Figure 4.1 Local-area DGNSS geometry.

moving. Relative GNSS thus cannot take advantage of near-perfect knowledge of the absolute location of the “master” receiver. However, the fact that most error sources discussed in Section 4.2.1 are highly correlated spatially means that, in a manner similar to DGNSS, the impacts of these errors are significantly reduced when relative GNSS positioning is performed (see [3]).

4.2.1 Error Sources and Degree of Spatial Correlation

Table 4.1 summarizes the principal error sources for users of the GPS Standard Positioning Service (SPS) [2], meaning “stand-alone” navigation based on L1 C/A-code pseudorange measurements (see [1, 3]). Where error sizes differ between pseudorange and carrier-phase measurements, “PR” in this table refers to pseudorange measurements, while “ ϕ ” refers to carrier-phase measurements.

Among the error sources listed in Table 4.1, all but multipath and receiver noise are strongly correlated spatially and are thus greatly reduced by the application of DGNSS. Only multipath is not affected, as it is statistically independent between reference and user receivers unless their antennas are very close together (i.e., $b \leq 20\text{--}30\text{m}$ in Figure 4.1). In fact, because of this independence, the use of DGNSS actually increases the total impact of multipath error because reference-receiver multipath (and receiver-noise) errors are added to those of the user. For this reason, it is important to minimize the impact of multipath on DGNSS reference receivers by careful siting and selection of receivers and antennas (e.g., see [4]). With reasonable care, this penalty is easily overcome by the error reduction achieved by the removal of well over 90% of, in rough order of magnitude, ionospheric, satellite clock, satellite ephemeris, and tropospheric errors [1, 3].

It is important to note the differences in spatial and temporal correlation among the first four error sources in Table 4.1 that are mitigated by DGNSS. All four of these error sources change relatively slowly over time, limiting the required update rate of the correction messages (see Section 4.2.3) and allowing slightly dated corrections to be projected forward in time and still be used, provided that the integrity time-to-alert requirements are still met (see Section 4.2.4 and Chapter 10). At a given time, satellite clock errors are perfectly correlated for all users and can be removed almost completely (the only limitation is the latency, or age, of

Table 4.1 Approximate 1-Sigma Errors for Stand-Alone and Local-Area DGPS Users

Error Source	<i>1σ Range Error for Stand-Alone Users</i>	<i>1σ Range Error for Local-Area DGPS Users</i>
SV clock error	1–2m	<0.05m
SV ephemeris error	1–3m (corrected by single-frequency model)	0.001–0.05m
Ionospheric delay error	1–7m	0.1–0.3m
Tropospheric delay error	2–3m (uncorrected) 0.1–0.5m (corrected by atmospheric model)	0.01–0.05m
Multipath (reference and user)	PR: 0.5–1.5m ϕ : 0.005–0.015m	PR: 0.5–2m ϕ : 0.005–0.02m
Receiver noise (reference and user)	PR: 0.2–0.35m ϕ : 0.001–0.003m	PR: 0.2–0.5m ϕ : 0.001–0.005m

the broadcast corrections—see Section 4.2.4). On the other hand, ephemeris errors, or errors in describing the true orbit of a GPS satellite, are not quite perfectly correlated because the projection of the satellite orbit error onto the line of sight from any receiver varies slightly with receiver location. The same is true of atmospheric errors (ionospheric and tropospheric delays) due to the fact that the atmospheric characteristics that generate signal delays are not precisely the same at different locations. The following paragraphs describe these three imperfectly correlated error sources in more detail.

The detailed analysis of residual DGNSS ephemeris error in [1] develops a bounding (worst-case) relationship between three-dimensional satellite orbit error (δR) for a given satellite and residual (post-DGPS) range error (ϵ_{eph}) as follows:

$$\epsilon_{eph} = \frac{\|b\| \delta R}{R} \quad (4.1)$$

where R is the distance between reference receiver and satellite (the distance from user receiver to satellite is essentially the same) and b is the baseline vector. For a typical local-area DGPS baseline, such as $\|b\| = 10$ km, and the smallest value of R ($\approx 20,180$ km), which occurs for a GPS satellite directly overhead, an ordinary three-dimensional satellite orbit error of 10m translates into a worst-case DGPS range error of only 0.005m. This is a worst-case result because it assumes the worst-case relative geometry between satellite, reference receiver, and user receiver (see [1, 5]).

During periods of non-negligible solar activity, ionosphere effects are now (in the absence of S/A) the largest source of error for stand-alone GPS users. Ionization within the upper atmosphere delays code-phase measurements and advances carrier-phase measurements by an equal amount, leading to errors of several meters or more in pseudorange measurements and code-carrier divergence in single-frequency code-carrier smoothing (see [3, 6]). The amount of ranging error that results varies significantly with the degree of solar activity and the elevation angle (θ) through which a given satellite is observed. Because this angle relates to the length of the signal that must pass through the ionosphere, satellites lower in the sky (i.e., smaller elevation angles) incur higher ionosphere delay than satellites higher in the sky. This effect is modeled by the so-called obliquity factor, which is derived based on approximating the ionosphere as a thin shell that concentrates its entire “delaying” impact at a single atmospheric altitude h_I . The resulting obliquity factor (OF) definition is [3, 7]:

$$OF = \left[1 - \left(\frac{R_E \cos \theta}{R_E + h_I} \right)^2 \right]^{-0.5} \quad (4.2)$$

where R_E is the mean radius of the Earth (6378.136 km), and the thin-shell height h_I is typically taken to be 350 km (a simpler approximation is defined in ICD-GPS-200C [8]).

The obliquity factor resulting from (4.2) ranges from 1.0 at 90 degrees elevation to just over 3.0 at 5 degrees elevation. Thus, satellites visible at 5 degrees elevation

will have approximately three times the ionospheric error as satellites visible very high in the sky. However, this approximation is only as good as the thin-shell model it is based upon. While that model appears to hold quite well under normal ionospheric conditions, significant deviations have been observed during ionospheric storms. These storms can cause severe spatial gradients that can magnify residual ionospheric errors for local-area DGPS users by a factor of 10 or more. Carrier smoothing of pseudorange errors, which is commonly applied by DGNSS users, magnifies this error because, as noted above, carrier-phase measurements are advanced by the same amount that pseudorange measurements are delayed (see [9, 28]).

Tropospheric errors are caused by signal delays due to the density of the neutral (noncharged) atmosphere within 10 km of the ground. The vast majority of tropospheric delay is denoted as the hydrostatic or “dry” component and is due to the refractivity of dry gases in the lower atmosphere. The remainder of the delay (roughly 10%) comes from the “wet” component and is due to the presence of water vapor in the lower atmosphere. The amount of water vapor in the atmosphere is highly variable; thus most of the short-term variation in the total tropospheric delay comes from the wet component (see [3, 9]). Tropospheric delays are significantly smaller than ionospheric delays under active solar conditions, vary much less over time, and affect code and carrier measurements identically.

The fact that tropospheric delays are generally well-behaved makes it relatively easy for all classes of GNSS users to remove most of their impact by using atmospheric models to predict the delay. These models typically take atmospheric pressure, temperature, and water-vapor partial-pressure data as inputs, but long-term average values for particular locations or regions can be used in place of real-time measurements with a relatively small loss of accuracy (see [3, 9, 10]). As with ionospheric error, a geometric factor $m(\theta)$ is needed to adjust for the user’s elevation angle. A very simple model for this is $1/\sin(\theta)$ [3], but models that are more accurate for small θ are preferred, such as the one given for use with the RTCA WAAS MOPS tropospheric model for $\theta \geq 4$ degrees [10]:

$$m(\theta) = \frac{1.001}{\sqrt{0.002001 + \sin^2 \theta}} \quad (4.3)$$

Tropospheric error correction models such as the ones cited in [9, 10] can be used by stand-alone users, but this is not a standard practice because tropospheric errors are not the dominant error source for standalone users. Wide-area DGNSS users (such as SBAS/WAAS users—see Chapter 10) do use these models to remove predicted tropospheric delays for their approximate locations. This makes sense because wide-area DGNSS users are spread over very large regions with widely different weather conditions, and wide-area “vector” corrections for different error sources are constructed in such a way that they do not include localized tropospheric effects to the extent possible (see Section 4.2.2). On the other hand, local-area scalar DGPS corrections include a very good representation of the tropospheric delay in a given locality; thus application of these corrections removes almost all tropospheric error from user measurements under normal weather conditions. The

most extreme weather conditions from the DGNSS user point-of-view would be the presence of sharp weather fronts and/or thunderstorms, which, for example, place the DGNSS reference station under cool, dry conditions while the user is under warm, humid conditions. A recent study of actual storms of this type (see [11, 12]) suggests that DGPS range errors of as large as 0.3–0.5m could occur over baselines as short as 5–15 km under worst-case conditions.

Having covered the four error sources that are removed by local-area DGNSS due to strong local spatial correlation, it is time to consider the effects of errors that are independent between user and reference receivers. Under nominal conditions, multipath is the largest error source for DGPS users because, rather than being canceled out by the application of DGPS corrections, it is enhanced somewhat by the combining of statistically decorrelated reference and user errors. The levels of multipath error shown in Table 4.1 assume “clean” or “moderate” reflecting-surface conditions around user and reference receivers and/or the use of at least some multipath-mitigating technology in the receivers or antennas (see [3, 4, 13]). Larger errors are possible under conditions with many or large reflecting surfaces nearby, as are sometimes encountered in urban areas (see [14]). Receiver noise is a function of received signal-to-noise power (see [3, 15]). User and reference receiver noise errors are normally reduced to relatively small values by carrier smoothing of C/A-code measurements, as described in [3]. The values shown for receiver noise errors in Table 4.1 assume that at least a default level of carrier smoothing (e.g., a time constant of 20 seconds or more in a first-order smoothing filter) is present. Longer time constants, such as the 100 seconds that is used in the Local Area Augmentation System (LAAS) (see Chapter 10) have the advantage of further reducing receiver noise and also attenuating high-frequency multipath errors.

Now that all major error sources have been considered, it is possible to estimate the total error budget for stand-alone and DGPS users under nominal conditions. Combining the error magnitudes listed in Table 4.1 and considering other small error sources that affect DGNSS users (such as the centimeter-level errors in knowledge of the reference-receiver location) suggests that typical one-sigma pseudorange errors for stand-alone GPS users are approximately 3–5m, compared to approximately 0.5–1.0m for local-area DGPS users. As mentioned before, because multipath dominates the local-area DGPS error budget under typical conditions (i.e., not considering the possibility of anomalies within one of the other error sources), further multipath mitigation beyond what is assumed in Table 4.1 has a direct effect in lowering total DGPS errors, whereas this is not the case for stand-alone users since other error sources not under the user’s control (e.g., ionospheric delays) are the dominant factors.

For applications such as surveying and precision farming that require centimeter-level accuracy, standard (pseudorange-based) local-area DGPS is insufficient. However, a variation of local-area DGNSS known as carrier-phase DGNSS (CDGNSS) [also known as real-time kinematic (RTK) GNSS positioning], routinely achieves decimeter-to-centimeter-level accuracy by providing carrier-phase in addition to code-phase corrections. The key to the use of CDGNSS is solving for carrier-phase integer ambiguities, which allows the very high-accuracy carrier-phase measurements (see Table 4.1) to be used as ranging observations. Details regarding

the algorithms that can be used to solve for carrier-phase integer ambiguities are discussed in Chapter 14 (also see [3, 14, 16]).

4.2.2 Local Versus Regional DGNSS Corrections and DGNSS Networks

Most of the discussion in this chapter refers to local-area DGNSS, which is the most common application of DGNSS. As shown in Figure 4.1, this application of DGNSS has one or more users applying corrections derived at and received from a single reference station, which contains one or more closely spaced reference receivers for the purposes of redundancy and measurement averaging. LAAS reference stations, for example, typically include four reference receivers with antennas spaced at distances on the order of 100m. Since corrections are generated from essentially one observing location, they are transmitted and applied as *scalar* corrections, meaning that one scalar value represents the pseudorange correction for a given satellite, and this single number covers all of the spatially-correlated error sources described in Section 4.2.1. This procedure works best when all users of these corrections are close to the reference station. As noted above, users within several tens of kilometers are best-served, but users as far as several hundred kilometers away can achieve substantial error reduction and safety improvement via local-area DGNSS, as evidenced by the success of the U.S. Coast Guard DGPS/radio beacon system fielded in the 1990s and since expanded into the “nationwide DGPS” (NDGPS) system, to support marine and other users (see [1, 17]).

Because the Coast Guard DGPS/radio beacon system employs many reference stations spread along the coastlines of the United States (with a few inland sites), it is an example of a DGNSS network. From a user perspective, the Coast Guard system operates just like single-reference-receiver local-area DGNSS because users normally apply only the corrections broadcast by the nearest (or easiest-to-receive) station. However, in other applications, measurements made by networks of reference receivers are used in combination to improve user performance and reliability, even though the resulting user corrections may appear to be from a single source. Integrated reference-receiver networks are better able to observe and correct for ionospheric and tropospheric spatial variations compared to what can be achieved at a single reference site. Regional networks that support CDGNSS are one common example of this approach (see Chapter 14 and [16]).

DGNSS systems that must support users spread over large regions, such as entire countries or continents, typically expand upon this “network” concept and make use of what is known as “wide-area” DGNSS. Unlike the single-reference-station system shown in Figure 4.1, or the NDGPS system with its many reference stations that each produce independent scalar corrections, wide-area systems combine the measurements of many reference stations spread out to cover the entire intended service area (or as much of it as possible, if some of it is oceanic). These measurements are combined at a so-called “master station” that generates a single set of correction messages applicable to users throughout the service area.

Broadcasting one set of corrections valid over a large region implies several key differences from local-area DGNSS. One is that communications between reference stations and the master station must be rock-solid. Also, while a well-built reference-station network can tolerate occasional outages of individual reference

stations, the master station must be sufficiently internally redundant that the probability of master-station outages is much smaller than the intended maximum reliability (or *availability*—see Chapter 10). Most fundamentally, the form of the corrections changes from scalar values for individual satellites to *vector* corrections: multiple values, which describe individual elements of the error sources that affect each possible user measurement.

The means by which corrections are implemented in space-based augmentation systems (SBASs), including the Wide-Area Augmentation System (WAAS) described further in Chapter 10, are a good example of the concept of vector DGNSS corrections. SBAS provides two separate sets of corrections, in two different formats, rather than the single scalar correction of local-area DGNSS. One set of SBAS message types broadcasts individual correction components for satellite-generated errors, including separate corrections for the one satellite clock error state and the three axes of satellite ephemeris error, for all GNSS satellites covered by a given SBAS at a given time. In addition, a “fast” correction message that combines clock and ephemeris corrections into one value for each satellite is broadcast at more frequent intervals to minimize latency (see Section 4.2.4) [10]. Meanwhile, a completely separate set of SBAS message types broadcasts ionospheric-delay corrections based on a grid of mostly 5×5 degree latitude/longitude points at the assumed ionosphere thin-shell height of 350 km (see [10, 18, 19]). SBAS users must follow the elaborate procedures laid out in [10] to combine these elements into measurement corrections applicable at their location and time (an example of this calculation is included in the software receiver presented in Chapter 5). The necessity for this can be seen from the fact that, in the case of WAAS, the same set of correction messages must be applicable to users in southern Florida, northern Minnesota, the Arizona desert, and off the west coast of the United States, even though these locations see different sets of satellites from different angles and see very different segments of the ionosphere.

Because wide-area DGPS corrections are generated to support large regions instead of being “specialized” for locations in the vicinity of a reference station, wide-area DGPS accuracy is significantly worse than local-area DGPS, although it is still much better than stand-alone GPS. The U.S. Federal Aviation Administration evaluates the performance of WAAS in quarterly “Performance Analysis Network” (PAN) reports, which are available online. A recent WAAS PAN report [20], covering January through March 2008, shows 95% (roughly 2σ) satellite clock/ephemeris errors of between 0.7 and 3.0m (most below 2m) after WAAS corrections are applied, depending on the (static) observing location and the satellite being observed, while 95% ionospheric residual errors after WAAS corrections range from 0.3 to 2.0m (most under 1m). The reference-to-user separation boundary, beyond which wide-area DGPS is likely to be more accurate than local-area DGPS, is estimated to be 200–250 km, but these figures underestimate the *integrity* advantages of wide-area DGPS, which are described in more detail in Section 4.3 and in Chapter 10 (also see [19]).

4.2.3 Means of Distributing DGNSS Corrections

Given the explosion in communication technologies that has occurred over the past 10–15 years, DGNSS corrections are now distributed over almost every

possible means of communication, including over the Internet (for example, see [21]) and over mobile-phone networks in support of A-GNSS technology (see [3, 22, 56]). However, most real-time DGNSS users continue to rely on some form of RF transmission.

Many means of transmitting DGNSS corrections via RF are described in [1]. One of these is the pre-existing marine radio beacon direction-finding technology that has been adapted to broadcast NDGPS signals over hundreds of kilometers. These radio beacons operate in the 283.5–325 kHz band and broadcast local-area corrections in the format defined by the Radio Technical Commission for Maritime Services (RTCM) Special Committee (SC) 104 (see [1, 23, 24]). This format is used by many local-area DGNSS systems around the world and has been adopted by the International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA). It consists of many different message types to transmit pseudorange corrections (PRCs), range-rate corrections (RRCs), reference station and satellite information, and a variety of other data types that may be useful depending on the characteristics of a particular DGNSS. Each message type is made up of a set of 30-bit “words,” each word being composed of 24 data bits and six parity bits. Furthermore, each message type begins with the same structure for the first two 30-bit words, and these two initial words define the broadcasting station identifier along with a “modified Z-count” that unambiguously gives the time that the message was transmitted (see [1]). It should be noted that the most recent published version of the RTCM-104 format, Version 3.1 [24], is not compatible with the latest version of the earlier format, Version 2.3 [23], as many changes were made in Version 3 to enhance CDGNSS applications. Version 3 also makes room for support of GLONASS, Galileo, and potentially other future GNSS constellations (see Chapter 17).

The original RTCM-104 data format, published in 1985, served as a model for the development of the subsequent Ground-Based Augmentation System (GBAS—see Chapter 10) data transmission formats. GBAS (including LAAS) correction information is broadcast at VHF in the band used for Instrument Landing System (ILS) localizer transmissions (108.0–117.975 MHz). This choice was made to allow aircraft to receive correction broadcasts with existing ILS avionics. Each 25-kHz interval from 108.025 to 117.950 MHz (inclusive) represents a usable center frequency, and since D8PSK modulation is used, each such center frequency has eight assignable time slots for broadcast messages [25]. One such slot is intended to be used by a single GBAS installation.

As with the RTCM SC-104 data format, the GBAS message structure encompasses several message types, of which Message Type 1 is the most important because it is updated twice per second and includes PRC, RRC, error bounds, and other integrity information. The other message types are updated less frequently. Message Type 2 contains information on the ground station location and on tropospheric and ionospheric error bounds, while Message Type 4 provides data that defines the approach supported by a given GBAS installation. Additional “optional” message types are defined to, for example, relay expected loss of satellite visibility (Message Type 5) [25].

While it is possible to distribute wide-area DGNSS corrections via terrestrial RF transmissions [the Ground-Based Regional Augmentation System (GRAS),

explained in Chapter 10, is an example of this], the best way to provide coverage of large regions is to broadcast corrections via satellite. As described in Chapter 10, SBAS systems (including WAAS) relay corrections and integrity information to users via satellites in geostationary (GEO) orbit (also see [10]). These messages, broken down into many different message types, are encoded as 250-bps navigation data streams on L1 C/A-code ranging signal broadcasts coming from each SBAS GEO satellite. Because encoding data on L1 C/A-code signals enforces a far lower data rate than the VHF broadcasts used by GBAS, yet SBAS is charged with providing vector corrections and error bounds to cover users located across vast regions, careful scheduling of individual SBAS message types is required to meet the six-second time-to-alert that applies to CAT I and “near-CAT-I” or “LPV” modes of precision approach (see Chapter 10).

4.2.4 Managing the Latency of DGNSS Corrections

While the lower data rate of SBAS corrections means that SBAS corrections suffer from much more latency than do GBAS corrections (or local-area corrections in general), all DGNSS systems need to adapt to the reality that users will receive and apply DGNSS corrections to their measured pseudoranges some time after the corrections were generated. In the meantime, the “zero-latency” correction, meaning the correction that applies at the current moment in time, has likely changed, if for no other reason than to keep up with changes in the satellite clock. It is for this reason that, at a minimum, RRCs are sent to users in addition to PRCs. With this information plus the time-tag t_C of the correction message (e.g., the modified Z-count in the RTCM and RTCA message formats) indicating when the PRC and RRC were generated, DGNSS users can easily compute a linear correction to update the PRC to apply to the user’s observation time t :

$$PRC_{upd} = PRC + RRC(t - t_C) \quad (4.4)$$

This approach is used by both GBAS and SBAS to remove first-order latency effects, but higher-order approaches may also be used (as in [26], which applies a quadratic time correction).

Whichever means of updating for latency is selected; the fact that users are applying dated reference-receiver information poses a potential vulnerability if GNSS satellite clock errors become much larger than normal and change dynamically in a manner that is not well-modeled by the latency-correction method. In the case of the linear correction applied in (4.4), if anomalous GNSS clock errors are limited to ramps with constant slope over time, these errors will be removed by the estimation of the ramp error magnitude in the RRC values broadcast by DGNSS reference stations. However, anomalous clock events limited to perfect ramps are unlikely; other changes, such as steps, accelerations, and jerks (acceleration rates) may also be present. In these cases, any practical latency correction is likely to leave significant residual error. While satellite clock faults are infrequent, they are among a class of anomalies that must be detected and excluded in order to insure user integrity at the 10^{-5} -per-operation level or better. Section 4.3 describes

how satellite faults and other anomalies can be mitigated by reference stations supporting DGNSS.

An additional consideration for latency that is related to both integrity and continuity is missed correction messages. If DGNSS users depend on reference stations to detect and exclude faulted measurements, missing one or more correction messages exposes them to potential faults that might have occurred since the time-tag of the most recent received correction message. For this reason, and because a time gap longer than a few seconds significantly degrades the accuracy of the linear latency correction given in (4.4), constraints apply regarding how many missed messages can be tolerated before operations must be aborted or downgraded in capability. The *time-to-alert* component of the integrity requirements for safety-critical users is typically the driving factor that determines how many consecutive missed messages are allowed (see Chapter 10 for more detailed discussion of this topic).

4.3 DGNSS Integrity Threats and Mitigations

One major advantage of DGNSS is the ability of reference stations to provide integrity to users in a manner that is more precise and less dependent on satellite geometry redundancy than is the primary alternative, which is known as receiver autonomous integrity monitoring (RAIM) (see [27–29]). Providing integrity for DGNSS users involves mitigating the following classes of faults and anomalies:

1. *Faults within GNSS*: This category includes failures within individual GNSS satellites and failures or errors in the control segment that manages a given GNSS constellation, such as the GPS Operational Control Segment (OCS).
2. *Faults within reference station equipment*: This category primarily includes failures within individual GNSS reference receivers as well as anomalous multipath at reference receiver antennas. It also includes potential faults in the communication methods used to relay DGNSS corrections and integrity information to users.
3. *Signal propagation anomalies*: This category includes anomalous spatial delay decorrelation when GNSS signals pass through the ionosphere or the troposphere on their way to reference and user antennas. The result of such anomalies is measured pseudorange delay at the reference receiver that is significantly different than the delay measured by users. Other impacts of anomalous propagation, such as ionosphere “scintillation,” are much more likely to cause loss of lock (or detection by one of the monitors looking for GNSS faults) and thus lead to continuity breaches (meaning unexpected loss of service) as opposed to loss of integrity (or a safety hazard—see Chapter 10).

Sections 4.3.1–4.3.3 cover each of these failure or anomaly categories separately.

4.3.1 Integrity Threats and GNSS Faults

This section addresses the five GNSS satellite and/or control segment faults defined by the FAA LAAS ground facility (LGF) specification [30]. This list includes all failure modes of significance to stand-alone or differential GNSS that have been seen, plus one that, to the authors' knowledge, has never been seen (satellite-generated code-carrier divergence). A brief description of each of these faults follows, along with a summary of the means by which DGNSS can mitigate the condition.

4.3.1.1 Satellite Code Signal Deformation

This fault mode involves the deformation of the code signal broadcast by a GNSS satellite relative to the ideal code signal expected by a GNSS receiver attempting to track it. Gross deformation of the signal would make the signal untrackable because receivers could not correlate to the flawed broadcast signal. These deformations do not constitute an integrity threat. However, more subtle deformations that retain the general shape of the desired signal can be tracked—they could distort the correlation peak as viewed by the receiver and thus result in hazardous errors. For DGNSS, if reference and user receivers track the faulted satellite identically, the resulting error will be identical and will cancel out once the user applies the reference-receiver-derived differential correction. However, the reference and user receivers are typically not identical; thus some differential error remains and could be of hazardous magnitude.

To the authors' knowledge, deformations with the potential to cause hazardous errors have only been observed once, on satellite SVN (and PRN) 19 in 1993 [31]. The fault appeared to cause a significant delay in the C/A-code rising edge chip transition relative to the P(Y)-code chip transition. After the fault was acknowledged by the GPS OCS and corrected by switching the satellite signal transmission electronics from the "A-side" to the "B-side," this anomalous transition delay was replaced with an acceptable level of delay. Note that, in practice, some level of deviation from "perfect" C/A-code exists on otherwise "nominal" satellites, and this "normal" signal deformation must be accounted for in the design of nominal error budgets and detection thresholds against anomalous, threatening signal deformations [31].

Figure 4.2 illustrates how the PRN-19 anomaly was first noticed by civil GPS users and demonstrates the vulnerability of differential users to it. The plot in Figure 4.2 is taken from differential GPS user measurements taken at the Oshkosh air show in August 1993. When PRN-19 was not in view, vertical errors were approximately 1–2m. When PRN-19 came into view at around 6:30 a.m. local time, vertical errors suddenly jumped as high as 8.5m (see [31] for a detailed study of this incident).

Concern over the effect of signal-modulation failures similar to PRN-19 led the aviation community to develop threat models to encompass the PRN-19 failure and reasonable extrapolations of it. Figure 4.3 illustrates the resulting "second-order-step" threat model that was created by RTCA and accepted by ICAO once the PRN-19 anomaly was better understood. This threat model includes the "digital delay (or advance)" fault mode thought to have been observed in the PRN-19 fault

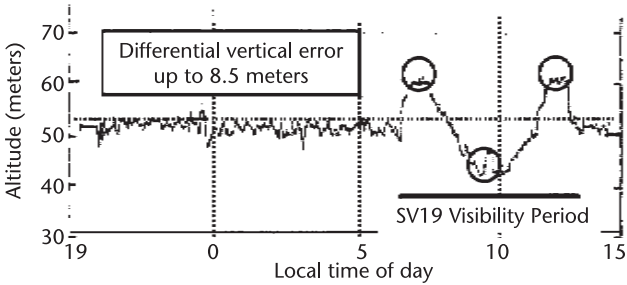


Figure 4.2 Vertical position error resulting from PRN-19 fault.

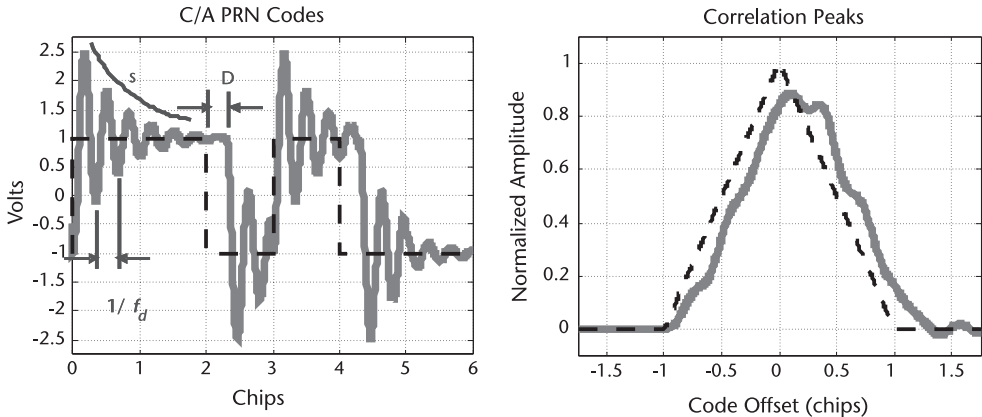


Figure 4.3 Threat model parameters and resulting correlation-peak deformation.

as “Threat Model A” with lead/lag parameter Δ . In addition, it includes a second-order-decay model of analog “ringing” during a code-chip transition as “Threat Model B” with parameters f_d (damped frequency) and σ (decay rate). “Threat Model C” combines both Threat Models A and B so that a single fault may have both digital-delay and analog-ringing characteristics. (This is the case illustrated by the thick, wavy line in Figure 4.3—see [7, 32].) The right-hand plot in Figure 4.3 shows the impact of the “Threat Model C” fault shown on the left-hand plot on the C/A-code correlation peak seen by a GPS receiver.

This class of signal deformation is capable of introducing appreciable errors in differential GPS systems. If the reference and user receivers use the same RF/IF transfer functions, the same discriminator scheme (e.g., early minus late and double delta), and the same correlator spacing, then the error suffered by the user will be the same as the error suffered by the reference. In this case, these errors would cancel in differential processing. However, if the reference and user differ with respect to any of these parameters (RF/IF transfer function, discriminator, or correlator spacing), then the errors will not cancel completely, and hazardous misleading information (HMI—see Chapter 10) may result.

Using the threat parameters described above, the deformed C/A-code signal (e.g., the thick, wavy line shown on the left-hand plot of Figure 4.3 relative to the ideal dashed line) is given by the following equations ([10, 33]):

$$e = \begin{cases} 0 & t \leq 0 \\ 1 - \exp(-\sigma t) \left[\cos \omega_d t + \frac{\sigma}{\omega_d} \sin \omega_d t \right] & t > 0 \end{cases} \quad (4.5)$$

$$\omega_d \equiv 2\pi f_d \quad (4.6)$$

Note that the signal-deformation case shown in Figure 4.3 represents the scenario where $f_d = 3$ MHz, $\sigma = 0.8$ MNepers/sec, and $\Delta = 0.3$. This case does not fall within any of the threat model boundaries defined below, as Figure 4.3 uses an exaggerated scenario to make the characteristics of deformation visually clear.

Table 4.2 gives the outer limits of the three threat-model parameters Δ , f_d , and σ as specified in two ways. The first, and primary, set of limits are from the “official” RTCA/ICAO threat model defined in [10, 30]. A secondary and tighter (meaning more constraining on the possible threat parameters) set of “most likely subset” (MLS) threat model parameters also is shown (see [33]). This second set is what the initial implementation of WAAS (approved for operation in July 2003) was subjected to. Both sets of limits are based on consensus-of-experts expansions from an estimate of the PRN-19 anomaly projected into the domain of the second-order-step threat model. For the most likely subset, the expansion from the PRN-19 case is very limited. This subset does not include the analog-only (Threat Model B) threat possibility. It does include the combined analog-digital threat (Threat Model C), but this threat is limited to a single frequency ($f_d = 10.23$ MHz).

In assessing the overall threat of signal deformation to DGNSS, a simulation is normally employed to search through a discretized set of points within and at the edges of one of these threat models to determine (1) the maximum differential user error in the range domain before any monitoring and (2) the maximum differential user error after taking credit for signal deformation monitoring (SDM) within its time-to-alert (e.g., 3 seconds for a CAT I LAAS Ground Facility—see [34] for a detailed treatment).

Figure 4.4 shows the basic concept behind signal deformation monitoring (SDM). This monitor makes code correlator measurements at multiple “early” and “late” chip offsets as well as at a “prompt” measurement. The monitor then forms detection statistics that check the correctness of the shape of the code correlation peak. These detection statistics can take the form of a collection of “ Δ -test” and “ratio test” statistics [32], each of which has its own threshold. These thresholds

Table 4.2 Signal Deformation Threat Model Parameter Bounds

	<i>Full RTCA/ICAO Threat Model</i>	<i>MLS Threat Model</i>
Threat Model A (Digital Lead/Lag Only)	$-0.12 \leq \Delta \leq 0.12$	$-0.04 \leq \Delta \leq 0.04$
Threat Model B (Analog “Ringing” Only)	$\Delta = 0$ (N/A) $4 \leq f_d \leq 17$ MHz $0.8 \leq \sigma \leq 8.8$ MNepers/sec	None (N/A)
Threat Model C (Digital Lead/Lag plus Analog “Ringing”)	$-0.12 \leq \Delta \leq 0.12$ $7.3 \leq f_d \leq 13$ MHz $0.8 \leq \sigma \leq 8.8$ MNepers/sec	$-0.04 \leq \Delta \leq 0.04$ $f_d = 10.23$ MHz $1.8 \leq \sigma \leq 7.8$ MNepers/sec

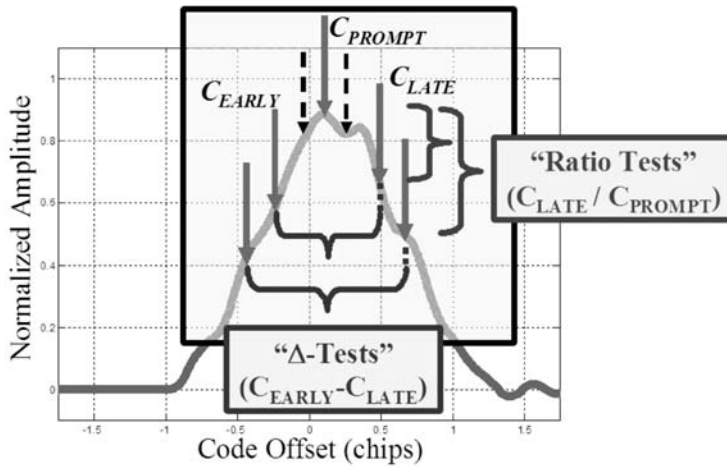


Figure 4.4 Illustration of multiple-correlator signal deformation monitoring.

are determined to provide an acceptably low level of false alarms under nominal conditions when the code correlation peak is influenced by nominal signal deformation (see [31]), receiver noise, and code multipath. In this approach, all Δ -tests and ratio tests must be passed at each epoch for the monitored signal to be declared healthy. More recent approaches “fuse” multiple test statistics into a single SDM test statistic [35, 36].

Note that, in the LAAS implementation, the raw correlation-peak observations shown in Figure 4.4 must be averaged or smoothed over a period similar to that of the 100-second carrier-smoothing time constant in order to reduce receiver noise and multipath down to acceptable levels. Therefore, careful analysis (and, in some cases, simulation) is required to demonstrate that all signal-deformation threats within the threat models given in Table 4.2 are mitigated within the time-to-alert if they generate potentially hazardous range-domain errors. The methodology for performing this analysis is detailed in [34], which introduces an approach called a time-varying maximum error range requirement (MERR). This approach has the useful property of relating the definition of the not-to-exceed error in the range domain (MERR) to the time-to-alert requirement. The “time-varying MERR” concept can be applied to all threats except for ionosphere anomalies, which are too large to be bounded in the range domain.

4.3.1.2 Satellite Low Signal Power

This fault involves the broadcast of satellite signals at power levels below the specified minimum levels. For GPS, the minimum received signal power levels are given in [8] (e.g., -160.0 dBW for C/A-code). In practice, GPS satellite broadcast signal power is and has been consistently several decibels above these levels. As reported in a study for the GPS Joint Program Office in 2002 [37], instances of significant violations of these minimum power levels are rarely (if ever) noticed, and our best understanding is that this continues to be the case today.

A threshold test on the measured C/N_0 in DGNSS reference receivers is described in [38] as a means of detecting low signal power (also see [15]). Through rare, signal power fades of a few decibels are important and should be detected. Deep fades (tens of decibels) do not constitute an integrity threat, because the reference and user receivers would simply lose lock. Shallow fades, on the other hand, could introduce cycle slips, and a few such slips could constitute an integrity threat. In addition, low satellite power could be indicative of other potential problems with the affected satellite. When an observable fault (such as low signal power) is detected on a given satellite, that satellite should be excluded from use, even if there is no immediate direct threat to integrity, because of the indirect threat to integrity posed by using satellites known to be in an anomalous state.

Low signal power can also appear as a consequence of local phenomena such as receiver failures, or anomalous reference-antenna multipath, and in these cases, direct integrity hazards are possible due to the effects of these other anomalies. In such “multisymptom” cases, the presence of low signal power in addition to other, more hazardous fault modes makes life easier for DGNSS integrity monitoring in that a separate monitor not targeted at the underlying fault (in this case, the C/N_0 monitor) might detect the unfolding anomaly before the “targeted” monitor does.

4.3.1.3 Satellite Code-Carrier Divergence

This is defined as a lack of synchronism between the code and carrier waveforms transmitted by the satellite. The effect is to cause a difference between the apparent range as indicated by code and carrier measurements. As measured by ground-based receivers, satellite-generated code-carrier divergence may differ from ionosphere-generated divergence. The ionosphere causes code delay and carrier advance to have equal magnitude (and opposite sign—thus “delay” and “advance”). Satellite code-carrier divergence refers to divergence of any type (not necessarily equal and opposite, and without an upper bound).

Unlike the other four satellite failures described in this section, satellite code-carrier divergence has never occurred, to the authors’ knowledge. In principle, it could be excluded based on fault modes and effects analysis (FMEA) directed at such failures in the satellite hardware. Since complete exclusion of such a fault based on FMEA analysis is very difficult in practice, and because no official guidance has been provided, code-carrier divergence is included in the list of possible failures because the effect of such a failure could definitely be hazardous. Most DGNSS systems (including SBAS and GBAS) use carrier-phase smoothing of code-phase measurements, and this process assumes that the code and carrier are synchronous when the signal leaves the satellite.

DGNSS systems that rely on single-frequency carrier smoothing in both reference and user receivers are vulnerable to this threat to the degree that their carrier-smoothing filters are not identical and have not achieved steady-state performance [39]. Monitoring for this threat involves estimating the degree of code-carrier divergence present in reference receiver measurements and comparing it to a threshold that is mostly based on what is typically present due to normal ionospheric divergence [38, 39]. Since experience suggests that satellite-generated code-carrier divergence is extremely unlikely, the combination of satellite clock monitoring

(described in the next section) and code-carrier divergence monitoring is sufficient to mitigate this potential hazard.

4.3.1.4 Satellite Clock Anomalies

Among the five classes of GPS satellite or control-segment faults listed above, satellite clock anomalies the most frequent and best understood. The cesium and rubidium atomic frequency standards that maintain long-term time stability in GPS satellites are known to unpredictably “lose control” as they age, despite careful monitoring and maintenance by the GPS OCS (see [40, 41]). The impact of this instability can be sudden changes in apparent range, range rate, and range acceleration as seen by reference and user receivers. Note that, unlike the code-carrier divergence fault mode defined above, both pseudorange and carrier-phase will move identically under the influence of satellite clock anomalies.

Two examples of satellite clock faults that occurred prior to 1997 are described in detail in [40]. Figure 4.5 shows a more recent clock failure that occurred on SVN/PRN-22 on July 28, 2001, as viewed by the FAA William J. Hughes Technical Center LAAS Test Prototype (LTP) installed at the Atlantic City, New Jersey, airport (see [42]). An initial large jump in clock drift rate was followed by a very fast clock runoff. Most worrisome, this fault was not flagged as “unhealthy” by the GPS OCS until one and a half hours after the onset of the anomaly.

Figures 4.6 and 4.7 provide two examples of less severe (and more typical) satellite clock failures from 2003 [35]. Figure 4.6 shows a clock drift problem on SVN/PRN-27 on May 26, 2003. This drift generated stand-alone user range errors

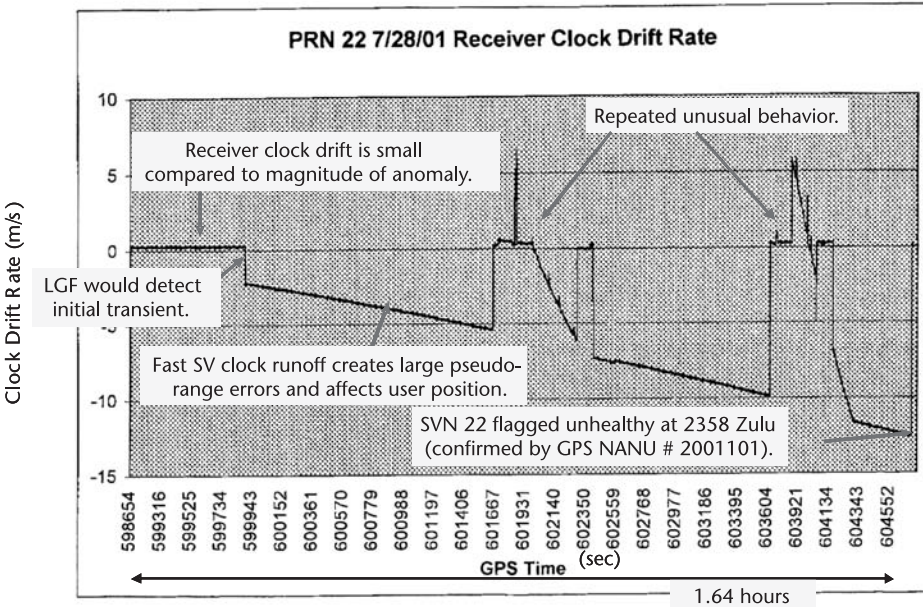


Figure 4.5 SVN/PRN-22 clock anomaly of July 28, 2001. (Figure courtesy of John Warburton, FAATC.)

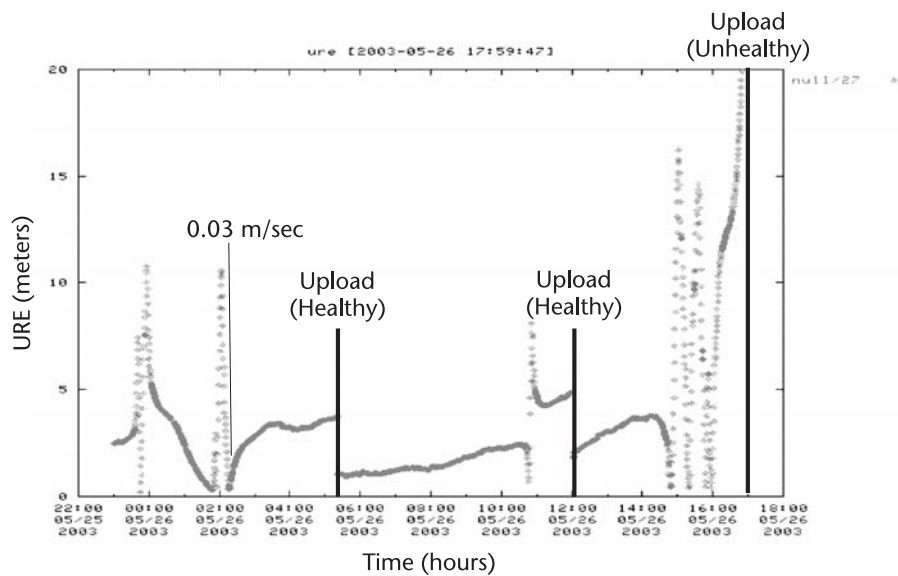


Figure 4.6 SVN/PRN-27 clock anomaly of May 26, 2003. (Figure courtesy of Karl Kovach for IFMEA.)

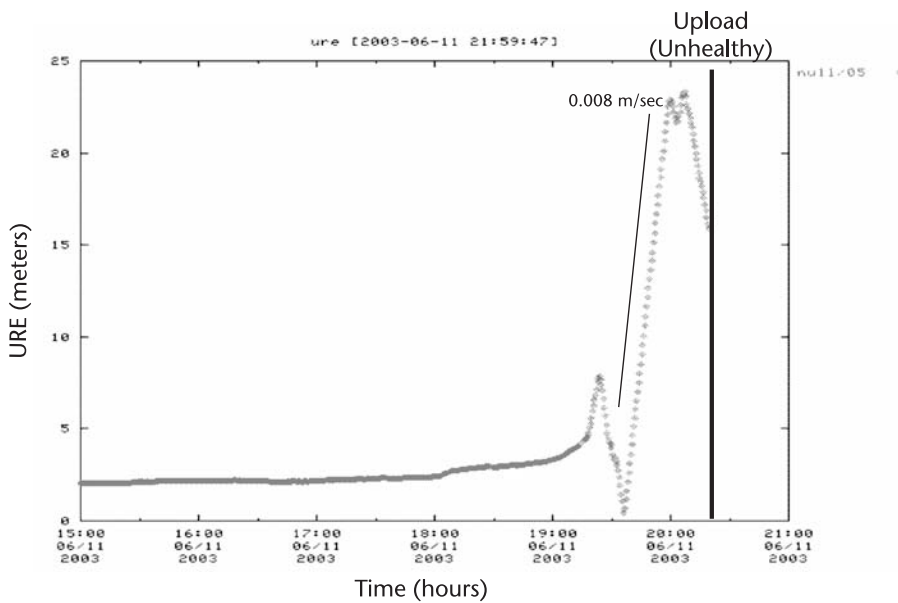


Figure 4.7 SVN-35/PRN-05 clock anomaly of June 11, 2003. (Figure courtesy of Karl Kovach for IFMEA.)

(URE)—shown on the y-axis—for OCS receivers that grew to 15m. Two attempts to correct it with updated satellite-clock-correction messages (in the broadcast satellite navigation data) failed, and the satellite was finally flagged as unhealthy after the URE shot upward and exceeded 20m. According to [40, 41], a URE of 20–30m represents the threshold that OCS uses to decide when a satellite should be flagged as “unhealthy” and taken off-line for further analysis before it is flagged

as “healthy” again. Additionally, the definition of HMI for stand-alone civil users in [2] is a URE from a satellite flagged as “healthy” that exceeds either 30m or 4.42 times the broadcast user range accuracy (URA) parameter. Figure 4.7 shows a simpler but otherwise similar example of clock “run-off” on SVN-35/PRN-05 on June 11, 2003. All three affected satellites in Figures 4.6 to 4.8 were GPS Block IIA satellites operating rubidium frequency standards.

These clock faults represent the most significant threat to stand-alone user integrity for two reasons. First, there is no clear upper bound to the size of possible errors in range, range-rate, and/or range-acceleration. Second, these faults occur with a relatively high frequency compared to other satellite fault modes, and the existing GPS OCS is unable to warn users of such faults within five to 10 minutes. Thus, stand-alone users requiring real-time integrity need to turn to one of several RAIM solutions, such as the one prescribed by RTCA DO-208 [36] for stand-alone aviation users (also see [28]).

For DGNSS users, these clock faults present a more complicated situation. Under ideal conditions, differential users would apply the reference corrections promptly and thus remove any impact of satellite range error changes. Since zero latency is impossible in practice, the degree to which differential users are directly threatened by clock anomalies depends on (1) the delay between reference correction generation and user application of the resulting corrections; and (2) the interpolation mechanism used to correct for this delay, as described in Section 4.2.4. Since the linear-range-rate latency correction given in (4.4) is the one most commonly used, satellite clock errors with significant range-acceleration error components are the most threatening to DGNSS because such accelerations make this method erroneous.

A means to detect the initial range-domain transient that results when a satellite clock becomes anomalous is described in [38]. This method is based on the carrier-

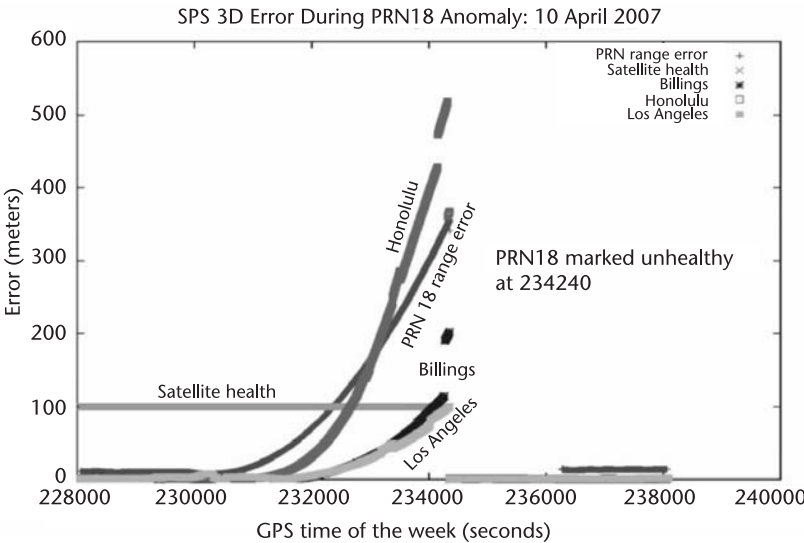


Figure 4.8 Errors resulting from SVN-54/PRN-18 ephemeris failure of April 10, 2007.

phase measurement extrapolation developed to mitigate latency in [26]. When a carrier-phase transient due to a satellite clock anomaly (or many other causes, including receiver cycle slips) occurs, the difference between the phase measurement predicted from previous measurement and the newly observed measurement (that is affected by the anomaly) provides a clear-cut means of detection.

4.3.1.5 Satellite Ephemeris Failures

These faults result in substantially erroneous ephemeris data being broadcast by a GNSS satellite. For GPS satellites, nominal ephemeris errors on a given orbit axis are less than 10m (1σ); thus a three-dimensional satellite orbit error exceeding 100m would certainly qualify as anomalous. Two very different causes of this type of anomaly have been foreseen:

- *Type A fault:* A large ephemeris error following a previously announced (A1) or unannounced (A2) satellite orbit maneuver;
- *Type B fault:* An OCS error in setting the broadcast ephemeris parameters (without a satellite orbit maneuver).

Prior to April 2007, type B faults were thought to create the larger integrity concern because they could occur due to OCS inattention when nothing unusual was happening. In contrast, type A faults were thought to be less likely because of the increased OCS vigilance associated with satellite maneuvers. In fact, no confirmed instance of any type of ephemeris fault as defined here had ever occurred until the recent (and serious) type A fault of April 10, 2007, which is illustrated in Figure 4.8. This plot comes from the SPS PAN report compiled by the FAA Technical Center for that period [43]. In this event, satellite SVN-54/PRN-18 was deliberately maneuvered from one orbit to another while still flagged as “healthy.” This serious blunder on the part of OCS should never have occurred, but it illustrates that all human and human-directed systems are fallible.

As shown in Figure 4.8, the apparent range error on PRN-18 during its unflagged maneuver grew to about 350m, meaning that the three-dimensional satellite orbit error was at least this large. Stand-alone GPS users in the western United States (here represented by stationary receiver sites maintained by the FAA Technical Center) saw three-dimensional position errors of several hundred meters depending on their location and the other (nominal) satellites in their position solutions. These errors persisted until the GPS OCS realized its mistake and set the satellite to unhealthy about 75 minutes after the “Delta-V” that triggered the planned orbit maneuver commenced. Note that the actual thrusting that triggered the orbit maneuver lasted only 44 seconds, after which the satellite was left to drift toward its newly desired orbit location.

As explained with regard to nominal ephemeris errors in Section 4.2.1, the impact of ephemeris anomalies on DGNSS users depends on two geometrical factors. First, as indicated by (4.1), the resulting differential range error increases linearly with reference-to-user separation ($\|b\|$). Second, the component of the three-dimensional satellite ephemeris error vector in space that results in differential range error is that component that projects into the plane that is perpendicular to

the vector connecting the reference receiver (or the geometric centroid between reference receivers) to the user. For example, if the reference-to-user vector (\mathbf{b}) points due west, only the component of the three-dimensional satellite ephemeris error (projected into east-north-up coordinates centered at the reference site) that falls within the north-south plane results in differential ranging error. The remainder of the ephemeris error is removed by the user's application of differential corrections.

Because there is no way to bound the magnitude of satellite ephemeris errors (particularly for type B faults) or to limit them to specific orbit axes, differential users require at least some degree of monitoring to reduce the ephemeris integrity risk to an acceptable level. This poses no problems for wide-area DGNSS because the regional spread of reference stations makes it possible to directly solve for satellite positions in space independently of the broadcast ephemeris messages (this is required to be able to generate vector corrections for individual ephemeris components) [44]. However, for local-area DGNSS, with reference receivers in only one location, satellite ephemeris monitoring is much more difficult.

Several monitors that are useful in detecting type B ephemeris faults within local-area DGNSS are described in [38, 45]. These methods are based on the "YE-TE" concept, which compares elements of "yesterday's ephemeris" (more precisely, the ephemeris stored from the most recent pass of a given satellite) to the same elements from "today's ephemeris," meaning the ephemeris that is broadcast by a newly risen satellite at the present time. The method described in [38] compares three-dimensional satellite positions in space, while the method in [45] compares several specific ephemeris orbit parameters and produces somewhat lower minimum detectable errors (MDEs).

Detection of type A anomalies, which involve actual satellite maneuvers, is even more difficult because the limited observability from a single reference location is combined with the lack of a valid prior ephemeris to serve as a reference. A differential-carrier-phase-based algorithm for detecting "Type A1" failures is described in [45]. This method attempts to make these maneuvers observable by utilizing the same ephemeris error decorrelation effect that poses a potential threat to users. Carrier-phase differences are used to make these effects visible over reference receiver antenna baselines of 100m to 1 km. In order to detect faults in all directions, two orthogonal baselines are needed within the set of reference receivers supporting nearby users.

Another, simpler approach to detecting type A anomalies is to apply thresholds on the differential range, range-rate, and range-acceleration corrections (note that the latter is not broadcast by LAAS) that are tighter than what is required to assure that the corrections fit within the broadcast message structure. For example, in LAAS, the maximum pseudorange correction that can be broadcast is $\pm 327.67\text{m}$ (see [25]), but a much smaller threshold of $\pm 125\text{m}$ can be applied to this correction without sacrificing continuity or availability. A threshold of $\pm 3.4\text{ m/s}$ can be applied to the range-rate correction, and a threshold of $\pm 0.019\text{ m/s}^2$ can be applied to the estimated range acceleration. These thresholds will detect almost all satellite maneuvers, whether commanded or uncommanded by OCS, but there is a small region of "worst-case" maneuver geometry that has little or no component along the vector between satellite and LGF that could go undetected and potentially (if

large enough) cause HMI. Thus, this approach does not preclude all possible threatening geometries. Its intent is to shrink the “space” of possible HMI events to the degree that the remaining space is of small enough probability that the overall integrity requirement is met (see Chapter 10).

4.3.2 Integrity Threats from DGNSS System Faults

This subsection focuses on DGNSS faults that directly affect receiver measurements. Software errors are not covered here because they are normally handled procedurally in high-integrity systems by the use of RTCA DO-178B [46] practices (not that these are perfect). Errors in the system that transmits corrections are also not covered. They can be managed by a combination of equipment redundancy and/or bit-to-bit comparison of the received broadcast to the data stream input to the transmitter.

4.3.2.1 Reference Receiver Errors and Cycle Slips

This category of errors covers everything that might go wrong inside a reference receiver that would cause it to produce incorrect pseudorange and carrier-phase measurements, which, if not detected and excluded, would corrupt the broadcast differential corrections. The inner workings of GNSS receivers are typically opaque to DGNSS designers. Hence, the implied “threat model” is conservative. It assumes that any magnitude of error in pseudorange or carrier phase is possible on any combination of pseudorange and/or carrier-phase measurements on one or more satellites being tracked. However, reference receiver failures are typically assumed to be statistically independent across reference receivers, meaning that the known presence of a fault on one receiver does not change the probability of a fault on any of the other receivers that appear to be working properly.

In practice, cycle slips are amongst the most common receiver faults, particularly when tracking satellites at low power levels [47]. In this fault, the carrier-phase measurement of the affected satellite(s) jumps by an integer (or half-integer) number of wavelengths of the carrier-phase waveform being tracked (e.g., 19 cm for L1). Usually, if cycle slips are large enough to be potentially hazardous, they are relatively easy to detect using the carrier-phase transient monitor described above for detection of satellite clock anomalies. In addition, if multiple redundant reference receivers are present, the multiple reference consistency check (MRCC) monitor used in LAAS detects faults of hazardous magnitude that are limited to a single reference receiver [30, 38]. MRCC uses B-values that represent differences among the redundant reference receivers to determine if a single reference receiver has failed on one or more of its channels.

Figure 4.9 uses simplified accounting to illustrate how B-values are computed in LAAS. Each reference receiver generates its own pseudorange correction (PRC_i) for a given satellite. Then, the corrections for all four reference receivers are averaged to produce the overall PRC that would be broadcast. The B-value for a given reference receiver j (B_j) represents the difference between this overall PRC and what the “correct” PRC would be if receiver j were known to be faulty. In this case, receiver j would be excluded from the generation of the broadcast corrections,

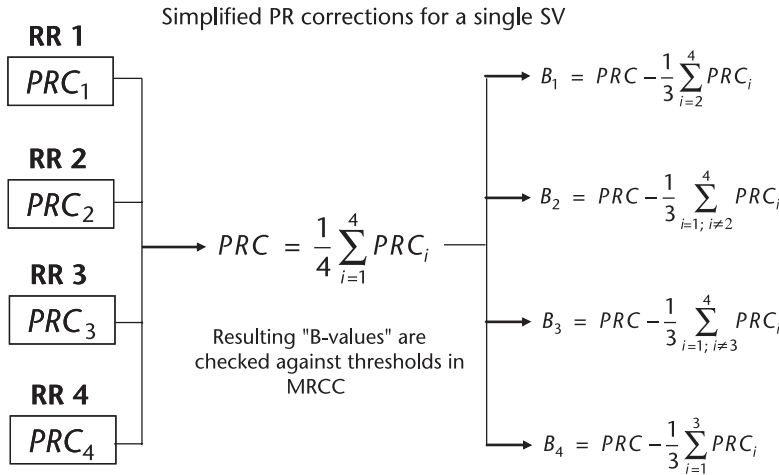


Figure 4.9 Illustration of simplified B-value calculation for LAAS.

and the “correct” PRC would instead be based on the average PRC over the three remaining (nonfailed) receivers. Each B-value thus expresses a measure of the difference between the receiver for which it applies and the consensus of the other three receivers, and if this difference exceeds a threshold, the receiver in question is marked for exclusion (at least with respect to the satellite(s) for which the discrepancy occurs). If one or more satellites fail MRCC, then a series of trial exclusions and recomputations of the remaining B-values is carried out until all remaining B-values fall below the threshold. At that point, the surviving B-values represent validated measurements that can be used to generate differential corrections (subject to passing all other monitors).

4.3.3 Integrity Threats from Signal Propagation Anomalies

These threats result from anomalies in the atmosphere between GNSS satellites and reference or user receivers. As shown in Table 4.1, under most conditions, ionospheric and tropospheric delays experienced by reference stations are almost identical (e.g., to within 10–30 cm) to those experienced by users within 100 km; thus, very little error remains after the application of differential corrections. Rare cases, with much greater differential delays across a 100-km baseline, create potential integrity hazards.

4.3.3.1 Ionospheric Spatial-Decorrelation Anomalies

Nominal one-sigma ionosphere vertical (zenith, referenced to a 90° elevation angle) delay gradients are from 1–4 mm/km in mid-latitude regions of the Earth. This finding is based on many years of experience with differential systems combined with extensive data analysis within the conterminous United States (CONUS) (see [48]). However, ionospheric delay gradients exceeding 400 mm/km have been observed on rare occasions during large geomagnetic storms. These gradients apply

to an arbitrary ray path and are not referenced to a particular elevation angle [49, 50].

Based on this data analysis, a LAAS ionosphere spatial anomaly threat model has been developed for CONUS to assess the impact of large ionosphere gradients on LAAS operations. The geometry of a large gradient is represented in Figure 4.10, where a linear ionosphere-slant-delay gradient is assumed to exist over a region of fixed width between regions of “flat” ionospheric delay (i.e., where the gradients are at nominal levels). This gradient “wave front” moves at constant velocity with respect to the ground.

In Figure 4.10, example values for the various threat-model parameters are shown that approximate a “near-worst-case” condition. Note that the front speed with respect to the LAAS Ground Facility (LGF) is 200 m/s, and the front is approaching from directly behind the aircraft so that the aircraft’s ionosphere pierce point (IPP) is impacted first. LAAS avionics built to the current MOPS [7] are not required to monitor for ionosphere anomalies; thus no help can be expected from user monitoring (even though such monitoring is possible—see [51]). Differential error begins growing before the wave front reaches the IPP, and when the LGF IPP is impacted, the LGF cannot detect the anomaly because the speed of the wave front is approximately the same as the speed of the LGF’s IPP. Thus, the code-carrier divergence (CCD) monitor described in Section 4.3.1 (to detect satellite-generated code-carrier divergence) is of no help in this ionospheric-anomaly scenario.

For single-frequency carrier smoothing, as used by LAAS and most local-area DGNSS systems, the differential pseudorange error (denoted here as ϵ_{iono}) resulting from the ionospheric anomaly modeled in Figure 4.10 is given by:

$$\epsilon_{iono} = g_{iono} (\|b_{air}\| + 2\tau\|v_{air}\|) \quad (4.7)$$

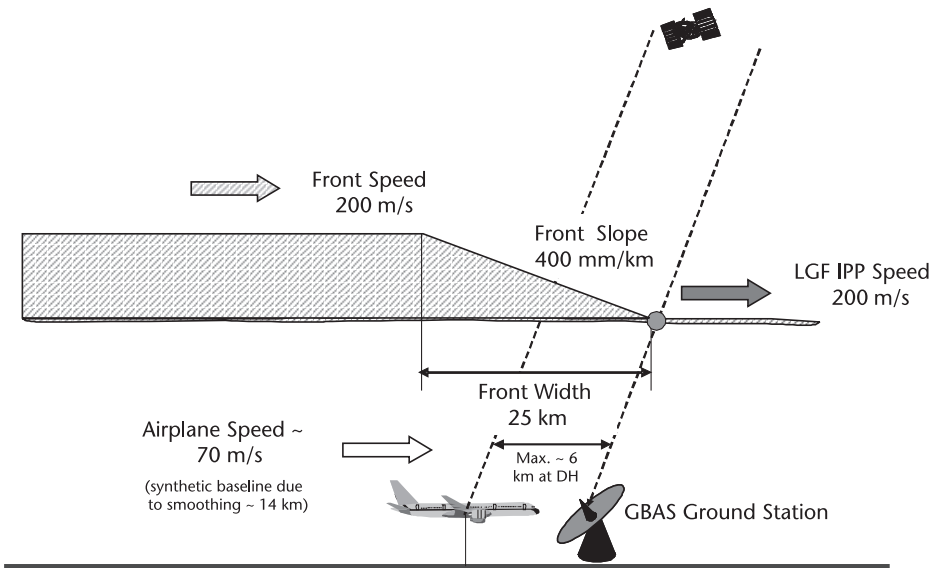


Figure 4.10 Geometry of near-worst-case ionospheric anomaly.

where g_{iono} is the ionosphere spatial gradient (in units of slope), $\|b_{air}\|$ is the baseline distance between LGF and aircraft, $\|v_{air}\|$ is the horizontal approach speed of the aircraft toward the LGF, and τ is the carrier-smoothing time constant (assumed to be the same in both LGF and aircraft, and typically 100 seconds for LAAS). A detailed derivation of this result can be found in [52]. Also note that the obliquity factor is not applied because anomalous gradients are already expressed in terms of slant delay. For the numbers shown in Figure 4.10, the resulting differential range error (on a single satellite) from (4.7) amounts to 8m. This is not disastrous for most DGNSS applications but could be hazardous for LAAS-supported precision approach and landing.

As noted above, extensive data analysis of ionosphere storms in CONUS since 2000 has allowed us to establish upper bounds for the threat-model parameters in Figure 4.10 (see [49, 50]). The plot in Figure 4.11 shows the upper bound on the ionosphere slant delay gradient (in mm/km) as a function of satellite elevation. This bound varies from 375 mm/km at low elevations to 425 mm/km at high elevations. These individual bounds express the largest gradients observed from ionospheric storms in CONUS. These largest gradients occurred in the Ohio-Michigan region during the extremely severe November 20, 2003, ionosphere storm between approximately 2030 and 2115 UT. These bounds also include a small margin (on the order of 10–20 mm/km) for measurement error. Also shown in the text within Figure 4.11 are upper and lower bounds on the width (25–200 km) and speed (0–750 m/s with respect to the ground) of the “wave-front” model shown in Figure 4.10 [49, 50].

To translate this ionosphere threat model into actual user impacts, a simulation methodology has been created that superposes the wave front shown in Figure 4.10 onto a given user satellite geometry. This procedure is described in detail in

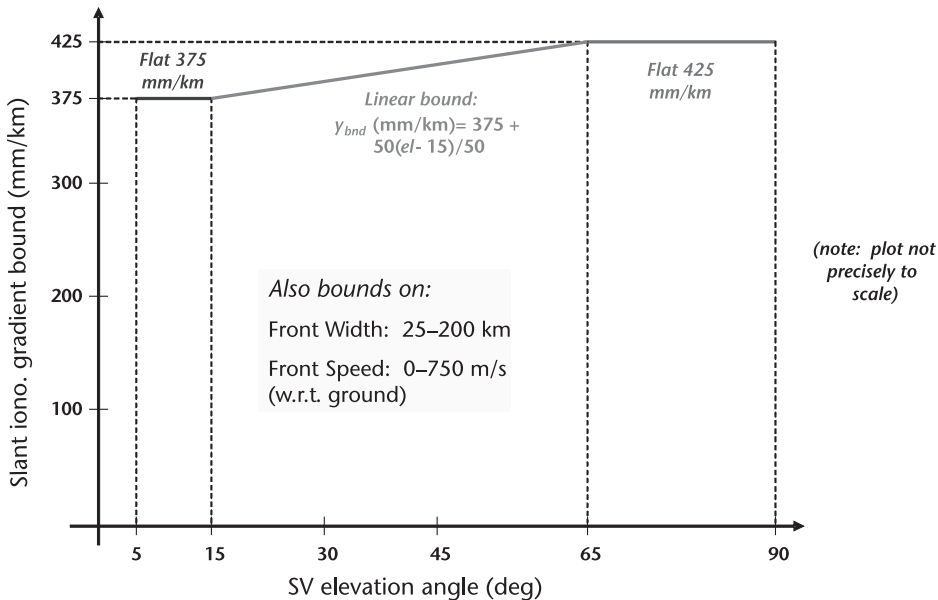


Figure 4.11 Ionospheric anomaly threat model parameter bounds.

[53]. The simulation applies either the worst-case ionospheric anomaly within the threat model or a randomly selected scenario within the threat model to either a single satellite or a pair of satellites in a given user geometry. The worst-case scenario for a given satellite is the maximum gradient for that satellite’s elevation angle, and it usually involves the minimum-width (25 km) front. It also includes the front speed and direction shown in Figure 4.10, where the aircraft is impacted first and the front speed nearly matches the LGF IPP speed for the satellite in question, making LGF detection impossible. For the two-satellite-impacted case, two cases are simulated. First, the worst-case anomaly is chosen for the first satellite in the pair, and the effect on the second satellite is then computed. Second, the worst-case anomaly is chosen for the second satellite and the resulting scenario for the first satellite is computed.

The worst-case vertical position error is called the maximum ionosphere error in vertical (MIEV) (position). The MIEV is found as follows. For each satellite and satellite pair in the subject geometry, the differential range error is converted into vertical position error using the known satellite geometry. The MIEV is the maximum position error over all satellites or pairs checked for that geometry. Importantly, the LGF (or the system operator) does not know the user’s exact satellite geometry. Hence, it must consider all possible user satellite geometries to determine the overall MIEV at a given location and time.

Figure 4.12 shows a histogram of vertical position errors for the Honeywell LAAS provably safe prototype (PSP) at Memphis airport for a CAT I approach with a decision height 6 km from the centroid of the LGF reference receivers. It is based on simulating a range of possible ionospheric anomaly scenarios within

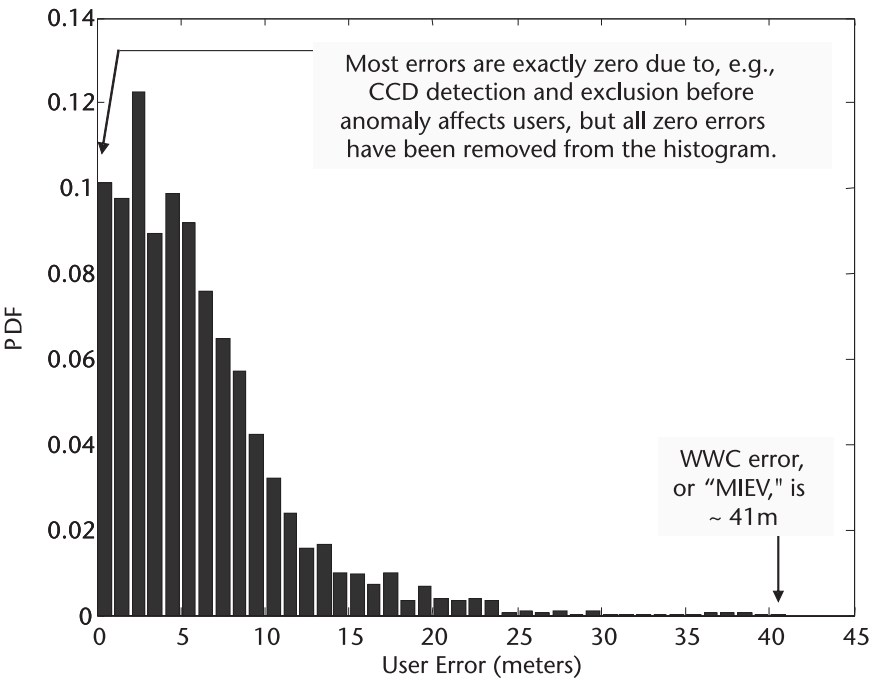


Figure 4.12 Histogram of vertical position errors for Memphis LAAS prototype.

the threat model. While all widths and velocities within the threat model of Figure 4.11 are simulated, only two maximum gradients are simulated (425 and 375 mm/km). Vertical position errors are computed for all satellites and pairs of satellites for the all-in-view user geometry, all “1-satellite-out” user geometries, and all “2-satellite out” user geometries. Simulating the threat model in this fashion results in most errors being identically zero because the LGF code-carrier-divergence monitor detection precedes the growth of any differential error, but these results have been deliberately excluded from the histogram. Even so, the vast majority of the resulting vertical errors are below 10m, although the very worst-case error (under the worst-case combination of ionospheric anomaly conditions, user geometry, and timing) is about 41m.

Several options exist for mitigating the risk of ionospheric spatial anomalies:

- Demonstrate that, for a given user class or set of applications, the probability of ionospheric anomalies severe enough to threaten user safety is acceptably low. If this can be done, no additional mitigation is required.
- Utilize signals on multiple GNSS frequencies to perform “divergence-free” and/or “ionosphere-free” smoothing, as described in detail in [54]. These techniques remove all or most of the differential ionospheric error between reference and user receivers and thus greatly reduce the magnitude of this threat.
- While only using a single GNSS frequency, add ionospheric-gradient monitoring to users as well as to reference receivers and broadcast information to users to allow them to “screen out” satellite geometries for which the worst-case ionosphere-induced positioning error would be hazardously large. This is the approach taken by a future upgrade of LAAS to support “GAST-D” CAT II/III service (see [51]).
- If none of the above are practical, the reference receivers must perform satellite geometry screening without the benefit of airborne monitoring and without knowledge of the exact user geometry. This is the case for the present version of LAAS supporting CAT I precision approach. In this method, the LGF determines the MIEV over all possible user geometries (typically down to two satellites removed from the set of satellites approved by the LGF), and if this MIEV exceeds the allowable worst-case error for this threat [55], the LGF must inflate one or more broadcast parameters that affect the airborne vertical protection level (VPL) such that all geometries with unacceptably high MIEV become *unavailable* [that is, VPL exceeds the tolerable vertical alert limit (VAL)] [7].

Two methods for implementing the technique proposed in the fourth bullet by real-time determination of inflation factors for combinations of the broadcast σ_{vig} , σ_{pr_gnd} , and/or the ephemeris P-value (see [25]) are given in [50, 53]. Because of the conservatism inherent in protecting against the worst possible ionospheric anomaly and the worst possible airborne satellite geometry, the resulting increase in VPL significantly degrades the achievable CAT I precision-approach availability using the existing “GAST-C” version of LAAS.

4.3.3.2 Tropospheric Spatial-Decorrelation Anomalies

As described in Section 4.2.1, tropospheric delays on GNSS signals are functions of local atmospheric pressure, temperature, and humidity. Thus, anomalous decorrelation of troposphere delays results from large differences in these weather parameters between the LGF and the user. As noted in that section, studies of tropospheric decorrelation for LAAS [11, 12] suggest that differential range errors as large as 0.3–0.5m can result from conditions where the LGF and the user are on different sides of a strong weather front (e.g., during severe thunderstorms) that separates warm, humid air from cold, dry air. While the resulting gradient is too small to be reliably detectable within GNSS measurements, the weather conditions under which such an event could occur are observable by other means (e.g., weather instruments, human observation of severe weather being present). Also, since this magnitude of worst-case differential error is an order of magnitude smaller than what is possible for ionospheric anomalies, special efforts to mitigate it are not required unless multiple-frequency measurements are available to mitigate ionospheric anomalies.

The approach recommended for LAAS in [12] is to include an independent one-sigma gradient of 5 mm/km to cover worst-case tropospheric spatial decorrelation. Since ionospheric and tropospheric anomalies are deemed independent, this 5 mm/km factor would be root-sum-squared (RSS'ed) with the 4 mm/km that is used to bound nominal (not anomalous) one-sigma ionospheric decorrelation in the broadcast σ_{vig} term (see [25, 48]), resulting in a “nominal” broadcast σ_{vig} (prior to any inflation to mitigate ionospheric anomalies) of 6.4 mm/km. This is unusual in that a sigma designed to bound worst-case errors is combined with one designed to bound nominal errors, but it is a conservative approach that requires no further mitigation of tropospheric anomalies.

4.4 Summary

This chapter describes the basic features of both local-area and wide-area differential augmentation of GNSS. DGNSS provides major performance improvements to satellite-navigation users over what is possible with stand-alone use of GNSS. In particular, DGNSS significantly improves both the *accuracy* and the *integrity* of user navigation. By removing error sources that are beyond the control of GNSS users, such as satellite and atmospheric errors, DGNSS moves the focus of error mitigation to error sources, such as multipath and receiver noise, that users (and DGNSS designers) have more control over. DGNSS reference receivers also serve as powerful means of real-time integrity monitoring that greatly exceeds the capabilities of a GNSS alone, even when augmented by user RAIM.

As new, improved GNSS constellations are fielded and existing systems continue to improve, the role of DGNSS may change. Much of the growth of GPS usage since the deactivation of selective availability in 2000 has been among personal and automotive users for whom the performance of stand-alone GPS is more than adequate. In the future, the presence of multiple constellations with higher basic signal accuracy will likely make it possible to support some levels of precision approach with stand-alone GNSS and RAIM (see [28, 29]), which would provide

(almost) worldwide service. However, for local and regional applications, the many new means of passing information from one place to another make it much easier to implement various forms of DGNSS “behind the scenes.” A-GNSS is an example of a use of DGNSS techniques to make GNSS signals easier to track in obscured locations [22, 56]. A-GNSS is also a means of meeting the ever-growing user expectations of GNSS performance. It is safe to say that DGNSS technologies, whether used openly for high-precision, high-integrity applications such as SBAS, GBAS, and surveying as done today, or covertly to boost the accuracy and coverage of next-generation automobile and mobile-phone navigators, will be standard elements of the GNSS designers’ toolbox for as long as GNSS remains in use.

4.5 Data Provided on the DVD

An example of DGNSS corrections has been provided on the DVD accompanying this book in the form of the WAAS option in the fastgps software receiver (to be presented in Chapter 5). Built into this receiver is the option to apply WAAS ionosphere differential corrections by using the 5-degree by 5-degree ionospheric grid point data described above (available at <http://www.nstb.tc.faa.gov/>). An example IGP file taken at the same time as software receiver data set 2 is included on the DVD, and it can be added into the processing loop using the configuration flag described in Chapter 5. The formulas for converting IGP grid data into corrections for individual pseudorange estimates can then be applied within the fastgps receiver (starting in the function `WAAS_corrections` in the file `nav.cpp`). This simple functionality is meant to provide a starting point from which others can expand the DGNSS functionality of the software receiver into more of the application areas discussed above.

References

- [1] Parkinson, B.W., and P.K. Enge, “Differential GPS,” in *Global Positioning System: Theory and Applications*, Volume II, B.W. Parkinson and J. J. Spilker (eds.), Chapter 1, pp. 3–50, Washington, D.C.: AIAA, 1996. <http://www.aiaa.org/content.cfm?pageid=360&id=568>.
- [2] *Global Positioning System Standard Positioning Service Performance Standard*, Washington, D.C., U.S. Department of Defense, Oct. 2001. <http://www.navcen.uscg.gov/gps/geninfo/2001SPSPPerformanceStandardFINAL.pdf>.
- [3] Misra, P., and P. Enge, *Global Positioning System: Signals, Measurements, and Performance* (second edition), Lincoln, MA: Ganga-Jamuna Press, 2006. <http://gpstextbook.com/>.
- [4] Lopez, A. R., “GPS Ground Station Antenna for Local Area Augmentation System, LAAS,” *Proc. ION 2000 National Technical Meeting*, Anaheim, CA, Jan. 26–28, 2000, pp. 738–742. http://www.ion.org/search/view_abstract.cfm?jp=p&idno=88.
- [5] Pullen, S., et al., “Ephemeris Protection Level Equations and Monitor Algorithms for GBAS,” *Proc. ION GPS 2001*, Salt Lake City, UT, Sept. 11–14, 2001, pp. 1738–1749. <http://waas.stanford.edu/~www/papers/gps/PDF/samion01.pdf>.
- [6] Klobuchar, J. A., “Ionospheric Effects on GPS,” in *Global Positioning System: Theory and Applications*, Volume I, Chapter 12, pp. 485–515, B.W. Parkinson and J. J. Spilker (eds.), Washington, D.C.: AIAA, 1996. <http://www.aiaa.org/content.cfm?pageid=360&id=568>.

- [7] *Minimum Operational Performance Standards for GPS Local Area Augmentation System Airborne Equipment*, Washington, D.C., RTCA, SC-159, DO-253B, June 26, 2007. <http://www.rtca.org>.
- [8] *Navstar GPS Space Segment/Navigation User Interfaces*, El Segundo, CA: ARINC Research Corporation, ICD-GPS-200C Rev. 4, April 12, 2000. <http://www.navcen.uscg.gov/pubs/gps/icd200/icd200cw1234.pdf>.
- [9] Spilker, J. J., "Tropospheric Effects on GPS," in *Global Positioning System: Theory and Applications*, Volume I, Chapter 13, pp. 517–546, B.W. Parkinson and J. J. Spilker (eds.), Washington, D.C.: AIAA, 1996. <http://www.aiaa.org/content.cfm?pageid=360&id=568>.
- [10] *Minimum Operational Performance Standards for Global Positioning System/Wide Area Augmentation System Airborne Equipment*, Washington, D.C., RTCA, SC-159, DO-229D, Dec. 13, 2006. <http://www.rtca.org>
- [11] Skidmore, T., and F. van Graas, "An Investigation of Tropospheric Errors on Differential GNSS Accuracy and Integrity," *Proc. ION GNSS 2004*, Long Beach, CA, Sept. 21–24, 2004, pp. 2752–2760. http://www.ion.org/search/view_abstract.cfm?jp=p&idno=5959.
- [12] Skidmore, T.A., and F. van Graas, "Accounting for Weather-Related Troposphere Errors in LAAS," Ohio University Avionics Engineering Center, Unpublished Presentation, Jan. 26, 2006. <http://www.ohio.edu/avionics/research/nav/gps/index.cfm>.
- [13] Braasch, M. S., "Multipath Effects," in *Global Positioning System: Theory and Applications*, Volume I, Chapter 14, pp. 547–568, B.W. Parkinson and J. J. Spilker, (eds.), Washington, D.C.: AIAA, 1996. <http://www.aiaa.org/content.cfm?pageid=360&id=568>.
- [14] Kubo, N., and S. Pullen, "Instantaneous RTK Positioning Based on User Velocity Measurements," *Proc. ION GNSS 2008*, Savannah, GA, Sept. 16–19, 2008. <http://www.ion.org/meetings/gnss2008/abstracts.cfm?track=E&session=3#p5>.
- [15] Van Dierendonck, A. J., "GPS Receivers," in *Global Positioning System: Theory and Applications*, Volume I, B.W. Parkinson and J.J. Spilker, (eds.), Chapter 8, pp. 329–407, Washington, D.C.: AIAA, 1996. <http://www.aiaa.org/content.cfm?pageid=360&id=568>.
- [16] Raquet, J., and G. Lachapelle, "Development and Testing of a Kinematic Carrier-Phase Ambiguity Resolution Method Using a Reference Receiver Network," *NAVIGATION: J. Inst. Navigation*, Vol. 46, No. 4, Winter 1999–2000, pp. 283–295. http://plan.geomatics.ualgary.ca/papers/Navigation_Raquetetal_1999.pdf.
- [17] Allen, L. W., III, "Nationwide Differential Global Positioning System," *Proc. 45th Civil GPS Service Interface Committee (CGSIC) Meeting*, Long Beach, CA, Sept. 12–13, 2005. <http://www.navcen.uscg.gov/cgsic/meetings/45thMeeting/09%20NDGPS,%20CGSIC,%20050912.ppt>.
- [18] Enge, P., et al, "Wide Area Augmentation of the Global Positioning System," *Proc. of the IEEE*, Vol. 84, No. 8, Aug. 1996, pp. 1063–1088. <http://ieeexplore.ieee.org/iel1/5/11069/00533954.pdf?tp=&arnumber=533954&isnumber=11069>.
- [19] Enge, P., and T. Walter, "The Wide Area Augmentation System," in *EGNOS: The European Geostationary Navigation Overlay System—A Cornerstone of Galileo*, J. Ventura-Traveset and D. Flament, (eds.), Chapter 4.1, pp. 395–411, Noordwijk, The Netherlands: The European Space Agency Publications Division, ESA SP-1303, Dec. 2006. http://www.esa.int/SPECIALS/ESA_Publications/SEMNMPLYE_0.html.
- [20] *Wide-Area Augmentation System Performance Analysis Report*, Atlantic City, NJ: FAA William J. Hughes Technical Center, NSTB/WAAS T&E Team, Report #24, April 2008. <http://www.nstb.tc.faa.gov/reports/waaspan24.pdf>
- [21] Soares, M. G., B. Malheiro, and F. J. Restivo, "An Internet DGPS Service for Precise Outdoor Navigation," *Proc. IEEE Emerging Technologies and Factory Automation Conference 2003 (ETFA '03)*, Lisbon, Portugal, Sept. 16–19, 2003, Vol. 1, pp. 512–518. <http://ieeexplore.ieee.org/iel5/8826/27939/01247750.pdf?tp=&isnumber=&arnumber=1247750>.

- [22] Enge, P., R. Fan, and A. Tiwari, "GPS Reference Network's New Role: Providing Continuity and Coverage," *GPS World*, Vol. 12, No. 7, July 2001, pp. 38–45. <http://www.gpsworld.com/>.
- [23] *RTCM Recommended Standards for Differential GNSS (Global Navigation Satellite Systems) Service*, Arlington, VA, RTCM SC-104, Version 2.3, RTCM 10402.3, 2001. <http://www.rtcn.org>.
- [24] *Differential GNSS (Global Navigation Satellite Systems) Service—Version 3*, Arlington, VA, RTCM SC-104, Version 3.1, RTCM 10403.1, Oct. 27, 2006. <http://www.rtcn.org>.
- [25] *GPS-Based Precision Approach Local Area Augmentation System (LAAS) Signal-in-Space Interface Control Document (ICD)*, Washington, D.C., RTCA, SC-159, DO-246C, April 7, 2005. <http://www.rtca.org>.
- [26] Lawrence, D. G., *Aircraft Landing Using GPS: Development and Evaluation of a Real Time System for Kinematic Position using the Global Positioning System*, Ph.D. dissertation, Stanford University, Dept. of Aeronautics and Astronautics, Sept. 1996. <http://waas.stanford.edu/~www/papers/gps/PDF/Thesis/DavidLawrenceThesis96.pdf>.
- [27] Brown, R. G., "Receiver Autonomous Integrity Monitoring," in *Global Positioning System: Theory and Applications*, Volume II, B.W. Parkinson and J. J. Spilker (eds.), Chapter 5, pp. 143–165, Washington, D.C.: AIAA, 1996. <http://www.aiaa.org/content.cfm?pageid=360&cid=568>.
- [28] Blanch, J., et al, "An Optimized Multiple Hypothesis RAIM Algorithm for Vertical Guidance," *Proc. ION GNSS 2007*, Fort Worth, TX, Sept. 25–28, 2007, pp. 2924–2933. <http://waas.stanford.edu/~www/papers/gps/PDF/BlanchIONGNSS07.pdf>.
- [29] Walter, T., et al, "Future Architectures to Provide Aviation Integrity," *Proc. ION NTM 2008*, San Diego, CA, Jan. 28–30, 2008, pp. 394–401. <http://waas.stanford.edu/~www/papers/gps/PDF/WalterIONNTM08.pdf>.
- [30] *Specification: Performance Type One Local Area Augmentation System Ground Facility*, Washington, D.C., U.S. Federal Aviation Administration, FAA-E-2937A, April 17, 2002. <http://gps.faa.gov/Library/Data/LAAS/LGF2937A.PDF>.
- [31] Mitelman, A., *Signal Quality Monitoring for GPS Augmentation Systems*, Ph.D. dissertation, Stanford University, Dept. of Aeronautics and Astronautics, Dec. 2004. <http://waas.stanford.edu/~www/papers/gps/PDF/Thesis/AlexanderMitelmanThesis04.pdf.zip>.
- [32] Phelts, R. E., *Multicorrelator Techniques for Robust Mitigation of Threats to GPS Signal Quality*, Ph.D. dissertation, Stanford University, Dept. of Aeronautics and Astronautics, June 2001. <http://waas.stanford.edu/~www/papers/gps/PDF/Thesis/EricPheltsThesis01.pdf>.
- [33] Shloss, P., et al, "A Simple Method of Signal Quality Monitoring for WAAS LNAV/VNAV," *Proc. ION GPS 2002*, Portland, OR., Sept. 24–27, 2002, pp. 800–808. <http://waas.stanford.edu/~www/papers/gps/PDF/PheltsIONGPS02.pdf>.
- [34] Rife, J., and R. E. Phelts, "Formulation of a Time-Varying Maximum Allowable Error for Ground-Based Augmentation Systems," *Proc. ION 2006 National Technical Meeting*, Monterey, CA, Jan. 18–20, 2006, pp. 441–453. <http://waas.stanford.edu/~www/papers/gps/PDF/RifeIONNTM06.pdf>.
- [35] Kovach, K., "Two Recent GPS Integrity Anomalies," El Segundo, CA: ARINC Engineering Services, LLC, Unpublished Presentation, June 25, 2003. http://www.arinc.com/products/spc_sat_eng/gps_engineering.html.
- [36] *Minimum Operational Performance Standards for Airborne Supplemental Navigation Equipment Using Global Positioning System (GPS) (incl. Change 1)*, Washington, D.C., RTCA, SC-159, DO-208, July 12, 1991 (Change 1: Sept. 21, 1993). <http://www.rtca.org>.
- [37] Edgar, C., D. B. Goldstein, and P. Bentley, "Current Constellation GPS Satellite Ground Received Signal Power Measurements," *Proc. ION 2002 National Technical Meeting*, San Diego, CA, Jan. 28–30, 2002, pp. 948–954. http://www.ion.org/search/view_abstract.cfm?jp=p&idno=287.

- [38] Xie, G., *Optimal On-Airport Monitoring of the Integrity of GPS-Based Landing Systems*, Ph.D. Dissertation, Stanford University, Dept. of Aeronautics and Astronautics, March 2004. <http://waas.stanford.edu/~www/papers/gps/PDF/Thesis/GangXieThesis04.pdf>.
- [39] Simili, D. V., and B. Pervan, "Code-Carrier Divergence Monitoring for the GPS Local Area Augmentation System," *Proc. IEEE/ION PLANS 2006*, San Diego, CA, April 25–27, 2006, pp. 483–493. http://www.ion.org/search/view_abstract.cfm?jp=p&idno=6677.
- [40] Crum, J. D., and R. T. Smetek, Jr., "Welcome to the Machine: An Overview of GPS Master Control Station Anomaly Detection and Resolution Techniques," *NAVIGATION: J. Inst. Navigation*, Vol. 44, No. 2, Summer 1997, pp. 133–152. http://www.ion.org/search/view_abstract.cfm?jp=j&idno=17.
- [41] Barker, B., and S. J. Huser, "Protect Yourself! Navigation Payload Anomalies and the Importance of Adhering to ICD-GPS-200," *Proc. ION GPS 1998*, Nashville, TN, Sept. 15–18, 1998, pp. 1843–1854. http://www.ion.org/search/view_abstract.cfm?jp=p&idno=3122.
- [42] Warburton, J., and D. Lamb, "Validation of the FAA LAAS Specification Using the LAAS Test Prototype (LTP)," *NAVIGATION: J. Inst. Navigation*, Vol. 45, No. 4, Winter 1998–99, pp. 265–274. http://www.ion.org/search/view_abstract.cfm?jp=j&idno=2235.
- [43] *Global Positioning System (GPS) Standard Positioning Service (SPS) Performance Analysis Report*, Atlantic City, N.J., FAA William J. Hughes Technical Center, NSTB/WAAS T&E Team, Report #58, July 31, 2007. http://www.nstb.tc.faa.gov/reports/pan58_0707.pdf.
- [44] Tsai, Y. J., *Wide Area Differential Operation of the Global Positioning System: Ephemeris and Clock Operations*, Ph.D. Dissertation, Stanford University, Dept. of Aeronautics and Astronautics, Aug. 1999. <http://waas.stanford.edu/~www/papers/gps/PDF/Thesis/YJTsaiThesis99.PDF>.
- [45] Pervan, B., and L. Gratton, "Orbit Ephemeris Monitors for Local Area Differential GPS," *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 41, No. 2, April 2005, pp. 449–460. <http://ieeexplore.ieee.org/iel5/7/31509/01468740.pdf?tp=&arnumber=1468740&isnumber=31509>.
- [46] *Software Considerations in Airborne Systems and Equipment Certification*, Washington, D.C., RTCA, SC-167, DO-178B, Dec. 1, 1992. <http://www.rtca.org>.
- [47] Schnaufer, B. A., and G. A. McGraw, "WAAS Receiver Carrier Tracking Loop and Data Demodulation Performance in the Presence of Wideband Interference," *NAVIGATION: J. Inst. Navigation*, Vol. 44, No. 1, Spring 1997, pp. 35–42. http://www.ion.org/search/view_abstract.cfm?jp=j&idno=10.
- [48] Lee, J., et al, "Assessment of Nominal Ionosphere Spatial Decorrelation for Global Positioning System-Based Aircraft Landing Systems," *AIAA J. of Aircraft*, Vol. 44, No. 5, Sept.–Oct. 2007, pp. 1662–1669. <http://www.aiaa.org/>.
- [49] Ene, A., et al, "A Comprehensive Ionosphere Storm Data Analysis Method to Support LAAS Threat Model Development," *Proc. ION NTM 2005*, San Diego, CA, Jan. 24–26, 2005, pp. 110–130. <http://waas.stanford.edu/~www/papers/gps/PDF/EneQiuIONNTM05.pdf>.
- [50] Ramakrishnan, S., et al, "Targeted Ephemeris Decorrelation Parameter Inflation for Improved LAAS Availability during Severe Ionosphere Anomalies," *Proc. ION NTM 2008*, San Diego, CA, Jan. 28–30, 2008, pp. 354–366. <http://waas.stanford.edu/~www/papers/gps/PDF/RamakrishnanIONNTM08.pdf>.
- [51] Murphy, T., and M. Harris, "More Ionosphere Anomaly Mitigation Considerations for Category II/III GBAS," *Proc. ION GNSS 2007*, Fort Worth, TX, Sept. 25–28, 2007, pp. 438–452. http://www.ion.org/search/view_abstract.cfm?jp=p&idno=7458.
- [52] Ko, P. Y., *GPS-Based Precision Approach and Landing Navigation*, Ph.D. dissertation, Stanford University, Dept. of Aeronautics and Astronautics, May 2000. <http://waas.stanford.edu/~www/papers/gps/PDF/Thesis/PingYaKoThesis00.pdf>.

- [53] Lee, J., et al, "Position-Domain Geometry Screening to Maximize LAAS Availability in the Presence of Ionosphere Anomalies," *Proc. ION GNSS 2006*, Fort Worth, TX, Sept. 26–29, 2006, pp. 393–408. <http://waas.stanford.edu/~www/papers/gps/PDF/LeeIONGNSS06.pdf>.
- [54] Konno, H., *Design of an Aircraft Landing System using Dual-Frequency GNSS*, Ph.D. Dissertation, Stanford University, Dept. of Aeronautics and Astronautics, Dec. 2007. <http://waas.stanford.edu/~www/papers/gps/PDF/Thesis/HiroKonnoThesis07.pdf>.
- [55] Shively, C. A., and R. Niles, "Safety Concepts for Mitigation of Ionospheric Anomaly Errors in GBAS," *Proc. ION NTM 2008*, San Diego, CA, Jan. 28–30, 2008, pp. 367–376. http://www.ion.org/search/view_abstract.cfm?jp=p&idno=7695.
- [56] Van Diggelen, F., *A-GPS: Assisted GPS, GNSS, and SBAS*, Norwood, MA: Artech House, 2009.

A GPS Software Receiver

Scott Gleason, Morgan Quigley, and Pieter Abbeel

5.1 Introduction and Background

This chapter presents a fully functional open-source software receiver. Using this GPS receiver, it is possible to demonstrate many of the basic navigation concepts introduced in the previous chapters, including acquisition, tracking, positioning, and applying differential corrections. The software receiver provided here will allow readers to experiment first-hand with the data sets provided on the accompanying CD, which contain real GPS signals. As a result, we hope, readers will derive a good understanding of how the various pieces of a typical GPS receiver operate.

The software receiver is known, rather ironically, as the **fastgps** software receiver. The fast in **fastgps** does not imply *fastest* but only that for a receiver processing everything in software, obtaining a position solution in less than a minute, as it is capable, is a notable feature. The receiver accomplishes this by being designed in C, using a number of speed optimizations, and having the option to use external aiding files that greatly reduce the amount of data processing needed.

The **fastgps** software receiver project was started by Morgan Quigley and Pieter Abbeel during their Ph.D. research in the computer science department at Stanford University. As a starting point in the receiver design, several ideas were taken from the open-source software-defined radio (SDR), written in MATLAB. The MATLAB software receiver was developed by the Danish GPS Center and is included in the book *A Software-Defined GPS and Galileo Receiver: A Single-Frequency Approach* by Kai Borre, Dennis Akos, Nicolaj Bertelsen, Peter Rinder, and Soren Holdt Jensen. The MATLAB SDR was written by Darius Plausinaitis, Kai Borre, and Dennis Akos. We would like to thank and acknowledge the authors of this work, for it provided a very useful reference during the initial development of the **fastgps** receiver. Over time, as the **fastgps** receiver advanced and other contributors became involved, it became a substantial original work in its own right with several new modules and features.

It is provided here with the hope that others will continue to make improvements and redistribute these advances openly to the research community. At the end of this chapter, we provide a list of suggestions for potential upgrades, from which we hope to entice interested GNSS programmers to become involved. Those wishing to get involved are encouraged to visit the **fastgps** project at SourceForge [1].

5.2 License, Development Environments, and Tools

5.2.1 License

The **fastgps** software receiver is released under the GNU general public license [2]. This is a valid legal copyright that ensures that others have the freedom to distribute this code with limited conditions, and without needing to obtain permission from the authors. However, it should be noted that the GPL license does require that the source code for any modifications be distributed back to the community, on the condition that the modified executable was released publicly. Additionally, some of the files included in the project are covered by different licenses and should these files be reused individually, the user is required to respect the existing copyright(s) for the code in question. These particular files are released under more flexible licenses that do not require the rerelease of modifications and are more suitable for incorporating in proprietary applications.

In summary, the receiver as a whole is covered by the GPL, while certain files within the receiver may be used in isolation according to the terms of their individual licenses.

5.2.2 GNU/Linux

The code was originally developed on a GNU/Linux PC using the freely available GCC GNU compiler tool suite [3]. For those interested in modifying the **fastgps** software we recommend a GNU/Linux or Mac OS environment. However, the code can be built and run on any platform and can use several publicly available compilers. Of particular interest on a GNU/Linux platform is the GNU profiler that can analyze the function-by-function performance of the code and pinpoint the areas in need of optimization as well as reveal hidden bugs.

Included on the DVD is a build environment that isolates the core library functions from the graphical user interface (GUI) build and the simpler console applications build. On the book DVD in the directory Chapter 5_SWRX/src there are three subdirectories. The fundamental receiver code, which does all of the real work, is included in the `libfastgps/` subdirectory. The code in this directory compiles into a library that is linked to the console build, found in the `fastgps/` directory. Alternatively, if the user is experienced with designing GUIs using wxWidgets [4], a (very) simple example is included in the `wxfastgps/` directory.

However, it is not necessary to have wxWidgets installed on one's PC to compile and run the receiver. The console version has all of the functionality of the GUI version, and the GUI code is meant to provide only a "foot in the door" for those wishing to add a little "bling" to the normal run environment. Prebuilt console executables for the three main platforms are included on the CD to allow users to get started in any environment they please by using the receiver configuration file to control most aspects of the receiver operation, as discussed in Section 5.4. To build and run the GUI versions it is necessary to install the wxWidgets libraries, which are included in the download repositories for most GNU/Linux distributions.

5.2.3 Microsoft Windows

As the core software modules are all written in standard ANSI C++, building and executing the receiver on a Windows platform is relatively straightforward with any number of compilers. The `fastgps` receiver has been tested using the MinGW GCC Windows compiler and the Borland free C++ compiler. Included on the CD are GUI and non-GUI makefiles for use on Windows with the MinGW and Borland compilers. They were generated using the cross-platform Makefile generator, `bakefile` [5]. Additionally, it would not be difficult to load the source code into any number of development environments, as the core source code is very portable.

5.2.4 Apple Mac OS X

The same Makefile used to build the GNU/Linux version of the receiver can be used on Apple Macintosh computers. Alternatively, the source code could be loaded into the Macintosh Xcode development environment [6] and run. As the Mac operating system is built around a GNU/Linux BSD variant called Darwin, the GNU/Linux Makefiles will all work in a Mac terminal window. Additionally, as the Mac OS development tools use a GCC compiler by default, and newer versions of Mac OS also come with a default version of wxWidgets GUI library, compiling and running the code is relatively simple. Prebuilt versions for both the GUI and console versions have been included on the DVD.

5.2.5 Displaying the Receiver Output

As the receiver runs it will generate several output files, which can be used to display the processing results at various stages. These outputs can all be controlled using the configuration file discussed in Section 5.4.1. As the formats of all these output files have been kept reasonably simple, it would be possible for users to easily write plotting routines using a range of available tools. However, to get things started, we have provided MATLAB/Octave and Python postprocessing scripts, in addition to the generated GoogleEarth navigation log, that are all capable of displaying the `fastgps` output results. Python [7] and Octave [8] were chosen because they are cross-platform and open-source applications that are well-suited for scientific analysis. Scripts with the extension “m” were designed to operate in both Octave and MATLAB without modification. Other tools that could easily be used for displaying and analyzing the receiver outputs include SciLab [9] and R [10], both of which are free and open-source.

5.3 Example Data Sets

Two example data sets have been included on the DVD, which can be processed by the `fastgps` receiver with or without the accompanying IGS ephemeris data file. The time and location of each data collection is described briefly in Sections 5.3.1 and 5.3.2.

5.3.1 Data Set 1

On May 15, 2008, an approximately 25-second data set was collected in Stoke Park, Guildford, United Kingdom, at nearly the exact center of a Cricket oval. The data was collected using a SiGe portable L1 data sampler [11] and contains the first three ephemeris subframes, enabling positioning without external aiding files. However, an IGS-aiding file (igu14794_06.sp3) is provided to reduce the time to the start of the navigation solution if desired. With the use of the aiding file the software can start estimating the receiver location after identifying only the first subframe, which occurs after the tenth second of sampled data. Figure 5.1(a, b) shows the logged position estimates using the Google Earth output file and the Octave/MATLAB plotting script.

5.3.2 Data Set 2, for Use with WAAS Corrections Data

The next day, on May 16, 2008, on the other side of the ocean, a longer data set was collected at the center of the Stanford University Quad in California. Shown in Figure 5.3 are the resulting positions generated by the **fastgps** receiver.

5.4 Using the **fastgps** Software Receiver

Most of the fundamental receiver inputs are controlled using a configuration file read at the start of execution. This permits the receiver to be used as is, without having to recompile it to make basic configuration changes. The configuration file can be changed to specify a range of input conditions, including the data set to be processed, the sampling rate and frequency, the logging and external aiding configurations (including self-aiding using its own output files), and the acquisition search strategy. Subsequently, the receiver generates several output files during the different phases of operation, depending on the logging flags included in the input file. These output files can be processed using the MATLAB/Octave and Python scripts provided.

Arguably, the most interesting way to analyze the receiver outputs is using the Google Earth kml output file. This file can be loaded directly into Google Earth and used to view the estimated positions in the context of satellite images of the collection area.

5.4.1 Configuration File

The receiver needs to know some basic information about the specified input data set in order to process the GPS signals captured in it. Additionally general flags on how the receiver should operate, what processing options it has, and how much information it should output are all best specified in an external text file that can be modified with any text editor and used to control some of the basic receiver operations. All of the input information is contained in the `fastgps_config.txt` file, which is formatted according to a couple of very simple rules.

Initially, the only character the parser looks for is “^.” The next character specifies which parameters will follow. Comments can be added, and existing

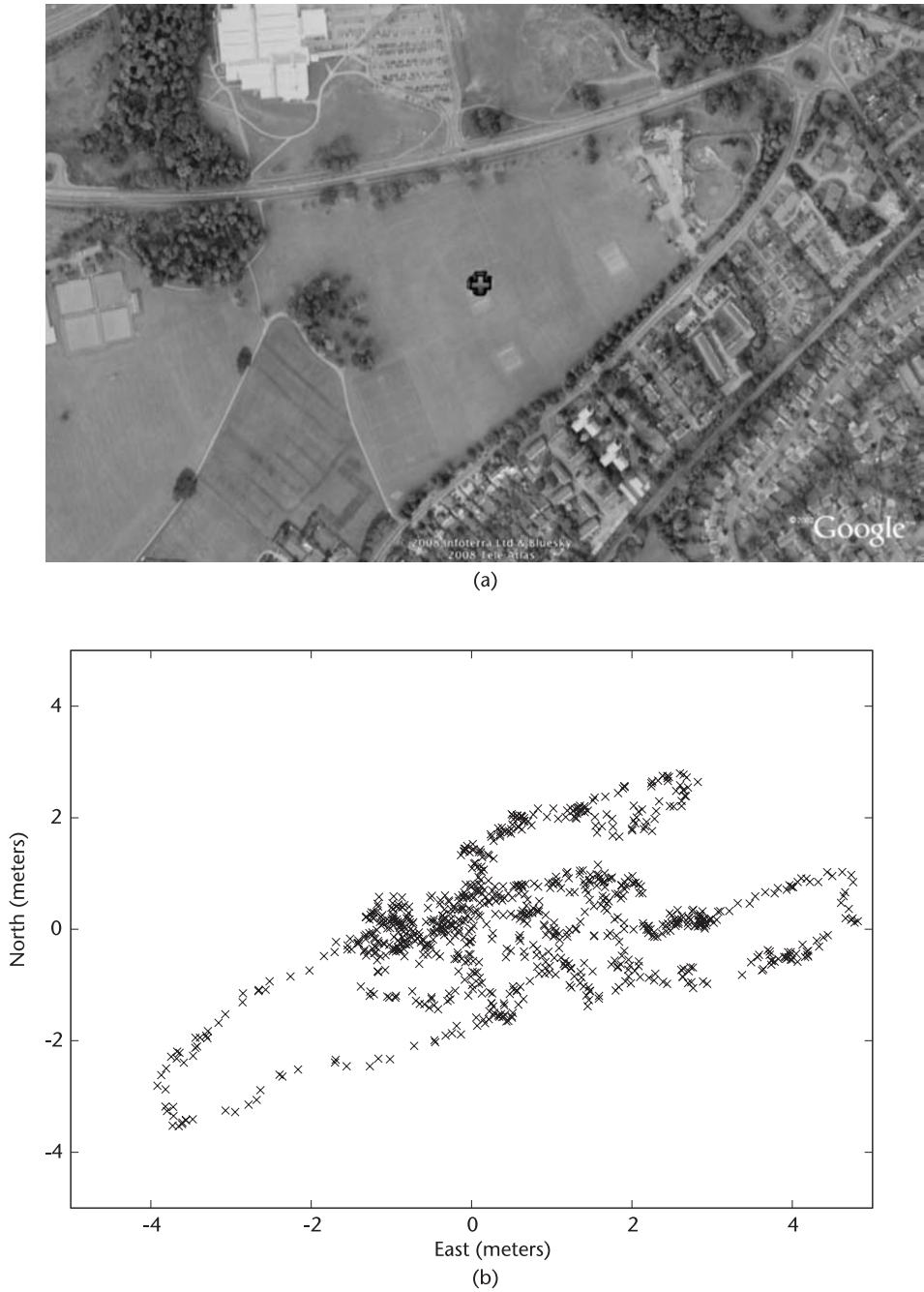


Figure 5.1 (a) The collection location of data set 1, as estimated by the **fastgps** software receiver, collected on May 15, 2008, in Stoke Park, Guildford, United Kingdom (background image provided by Google Earth) and (b) the east/north offset from the first navigation solution, plotted using the output scripts.

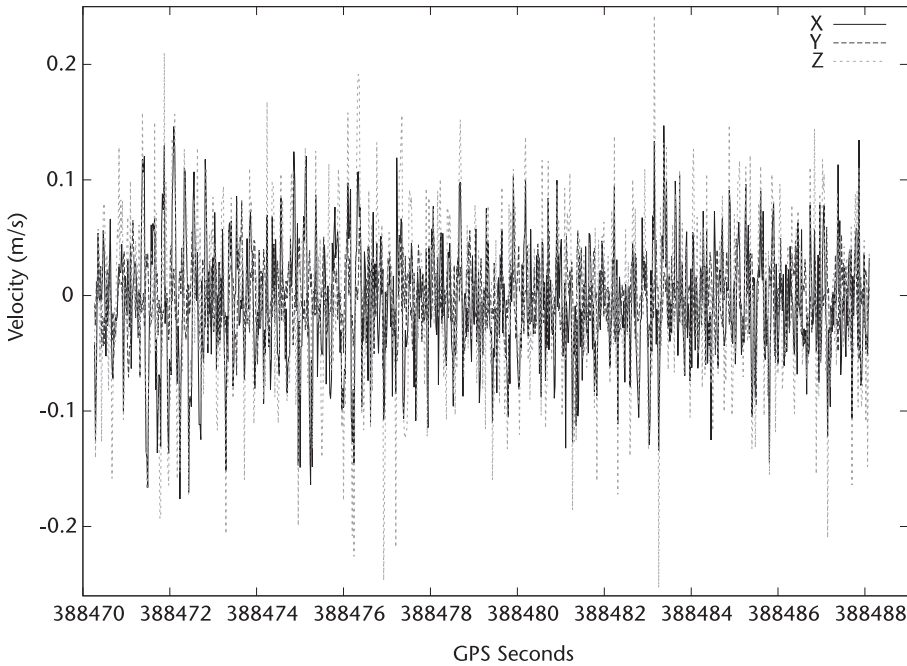


Figure 5.2 The receiver velocity estimates for data set 1, as determined by the **fastgps** software receiver, collected on May 15, 2008, in Stoke Park, Guildford, United Kingdom.

command lines can be “commented out,” by simply removing the caret from the beginning of a line. The command character directly after the caret lets the parser know the exact format of the parameters that will follow. Table 5.1 lists the available command characters.

An explanation of the available command characters follows:

- **Data file ('D')**: This entry is used to specify the raw data file to process. This can be a relative or an absolute path (within reasonable length). If the receiver cannot find the file an error message will be displayed and the processing stopped. The double backslashes are required on Windows operating systems.
- **Frequencies ('F')**: This is required to let the receiver know how the raw data file specified above was sampled by the RF front end. The first entry is the sampling frequency and the second is the intermediate frequency (IF), both in hertz. The IF is the frequency to which the raw signals (as seen by the antenna) are down-converted before sampling.
- **Satellite search strategy ('S')**: This setting is used to control how the receiver searches for satellites during the acquisition phase of operation. If the first parameter is 99, this tells the receiver to do a complete cold search for PRNs 1 through 31 and to use all that are found (up to 12). If the first parameter is 100, the receiver uses the existing acquisition log file to reacquire exactly the satellites it found previously. This can be used to skip the acquisition search and jump straight to signal tracking using an old acquisition file.

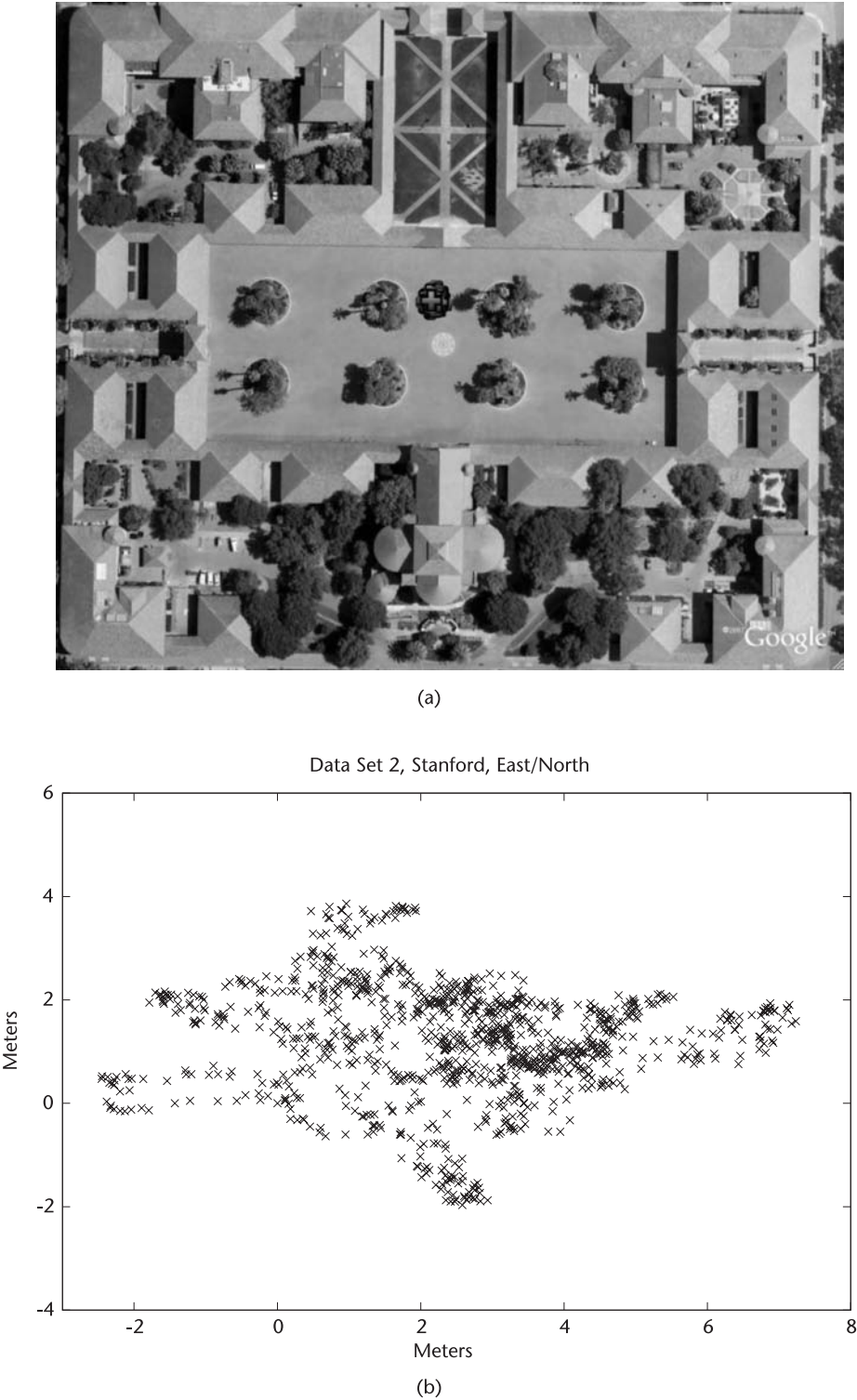


Figure 5.3 (a) The collection location of data set 2, as estimated by the **fastgps** software receiver, collected on May 16, 2008, in Stanford, California (background image provided by Google Earth), and (b) the east/north offset from the first navigation solution, plotted using the output scripts.

Table 5.1 fastgps Configuration File Command Summary

<i>Start</i>	<i>Command</i>	<i>Parameters</i>	<i>Example</i>
‘^’	‘D’	Data file w/path	^D \\data\\rawdata.dat # Windows ^D ../data/rawdata.dat # GNU/Linux, MacOS
‘^’	‘F’	Sampling freq, IF	^F 16.3676e6 4.1304e6
‘^’	‘S’	Satellite Search	^S 1 3 # search for 1 satellite, prn 3 ^S 2 3 18 # 2 satellites, prn 3 and prn 18 ^S 99 # all satellites, cold search ^S 100 # use acquisition log file to acquire
‘^’	‘T’	Processing time	^T 14.5 # process 14.5 seconds of data
‘^’	‘L’	Logging output	^L 1 1 1 # normal acq, tracking and nav logs ^L 2 1 3 # normal and debug acquisition logs, normal tracking log, normal and GoogleEarth navigation logs.
‘^’	‘G’	Tracking loop gains, 9 entries	^G 500 1000 1500 0.004 1.0 1.15 0.10 0.10 0.93 # default values
‘^’	‘E’	Use ephemeris file	^E 1 # use ephemeris log file
‘^’	‘A’	External Aiding on/off, week, day, sp3a filename	^A 1 1463 4 IGS_Data\\igr14634a.sp3 # XP ^A 1 1463 4 IGS_data/igr14634a.sp3 # Linux
‘^’	‘W’	WAAS Ionosphere corrections	^W ../WAAS_data\\IGP_delays.txt # XP ^W ../WAAS_data/IGP_delays.txt #Linux/Mac

Note: Bold entries required for all receiver runs.

Otherwise, the first parameter specified is the number of PRNs to follow, all of which will be searched for and tracked if found.

- Processing time (‘T’): This entry lets the receiver know how much data to process, in seconds. To do positioning, the length of data specified must include at least one subframe, assuming aiding data is activated. For positioning without aiding, the first three subframes are needed, a condition that can only be guaranteed with a data set 36-seconds-long (though it can be less with some luck). When one is only interested in acquisition or short tracking intervals, this parameter can be set to one second or less to allow quick repeatable runs.
- Logging output (‘L’): This represents a list of flags for each of the acquisition, tracking, and navigation functions, respectively. In all cases a 1 indicates normal logging, 2 normal + debug logging, and in the case of the navigation flag, a value of 3 activates both normal logging and the Google Earth output file.

For acquisition, normal logging consists of a simple file containing the basic acquisition information for all satellites found. This file can also be reloaded to quickly reacquire known satellites, using the “^S 100” option. Debug logging generates two additional files, one that contains the inverse FFT outputs for all satellites searched for, the other the results of a complete code delay search for each satellite at its estimated Doppler frequency. For the debug file only, the results for one satellite are included in each row of raw values. For tracking, normal logging stores several tracking loop outputs for all channels at an interval of 20 ms. If debug logging is specified the information remains the same but the log is updated every 1 ms. For naviga-

tion, normal logging stores raw position, velocity, and time estimates at the navigation output frequency (specified by the `#define PVT_INTERVAL` in `parameters.h`, with a default value of 20 ms) as well as raw pseudorange measurements and satellite information. Setting the navigation module flag to 3 generates an additional Google Earth-compatible kml file for displaying the estimated positions.

- Tracking loop gains ('G'): These are optional flags that control the tracking loop switching intervals (i.e., the seconds into processing when the FLL loop gains are changed and when the PLL is applied) and the corresponding gains for all the allocated tracking channels. Changing these codes is not recommended unless the user is experienced in GPS code and frequency tracking. However, the control loop design is described briefly in Section 5.5.4 and based roughly on the derivations provided by Groves [12], for those who wish to optimize or experiment with the signal tracking coefficients.
- Ephemeris file ('E'): This parameter specifies an optional flag that tells the receiver to load satellite ephemeris data from a file. This is useful when an external aiding file is not available. In this case, the ephemeris subframes can be decoded once during an initial processing run and then reused quickly on subsequent runs of the same data. Thus, on repeat runs on the same data set, it is not necessary to wait for the ephemeris to be decoded, for it is already loaded and positioning can start upon reception of the first subframe. If the ephemeris has not been decoded and saved for a data set, this flag should not be set.
- Aiding data ('A'): This entry indicates that an external aiding data file be used in place of the decoded satellite ephemeris information. This entry should not be included if the ^E entry above is activated. These files are publicly available for download from the International GNSS Service (IGS) [13]. The first parameter in this entry is a flag telling the receiver to use the external data specified, while the second and third parameters are the GPS week number and day (0 through 6) on which the data was collected (these must agree with and are included as part of the IGS filename). The routines for interpolating the IGS data were generously provided and upgraded specifically for the **fastgps** receiver thanks to Steve Hilla and others at NOAA. The interpolation routine is based on elements of the orbit prediction research performed by Mark Schenewerk [14].
- WAAS ionosphere corrections ('W'): This entry specifies that WAAS ionosphere differential corrections are available in a specified file. These files can be obtained from the FAA WAAS web site [15] and are included under the menu Real-Time Data → IGP Delays. The data obtained here can be copied into a text file and read by the software receiver to apply differential ionosphere corrections to individual pseudorange measurements. A brief summary of these and other WAAS corrections can be found in [16], while the detailed documentation of all WAAS packets is found in RTCA DO-229D [17]. For detailed information on GNSS differential corrections, including WAAS, see Chapter 4. An example input file is included for the second example data set on the DVD.

5.4.2 Output Files

The output files described below can all be processed with the provided MATLAB/Octave (`Process_fastgps_Outputs.m`) and Python (`Process_fastgps_Outputs.py`) scripts. These two scripts provide a simple interactive command line tool for plotting the information included in the generated log files. Alternatively, as the file formats (with the exception of the normal acquisition file) have all been designed as loadable arrays, writing a custom script using any number of applications should be relatively straightforward. Each of the output files generated by the **fastgps** receiver is detailed as follows.

- Acquisition (normal output) (`fast_acquisition_log.dat`): This file includes a slight twist, the addition of a 'A' or 'N' as the first entry. This was added to make it usable as a self-aiding input file for the receiver during initial acquisition (using the “^S 100” search option in the configuration file). The catch is that this character will confuse the MATLAB/Octave and Python interpreters if one attempts to load the file as an array. To overcome this, a sister file is also written (`fast_acquisition_log_plot.dat`) that includes a zero placeholder instead of the acquisition characters. This second file is loaded into the plotting routines, while the first can be used to initialize the receiver acquisition. Each of the 32 searchable satellite PRNs is represented by one row, consisting of the entries listed in Table 5.2.
- Acquisition (debug outputs): When the receiver is configured to output debug acquisition information, two additional files are generated: `acq_debug_log.dat` and `acq_debug2_log.dat`. The first file contains a row of raw results from the inverse FFT, taken over the entire 4-ms acquisition data block. For each PRN searched for, an additional row is added. Likewise for the second file, but this file contains raw outputs over all code phases

Table 5.2 Format of the Normal Acquisition Log File

<i>Column</i>	<i>Value</i>	<i>Comment</i>
0 (acq file)	'A' for an acquired satellite or 'N' if not acquired.	Necessary for input parser
0 (plot file)	0	(Placeholder)
1	Process time	Seconds
2	Process counts	Number of iterations of the main loop (see architecture below)
3	PRN number	1 through 32
4	Channel number	Channel (0 through 11) the PRN has been allocated to, if found
5	spare	
6	Max acquisition ratio	Max value/average value, if over the defined detection threshold
7	spare	
8	Carrier phase increment	Used for reacquisition
9	Code phase increment	Used for reacquisition
10	Carrier frequency	Used for reacquisition
11	Code frequency	Used for reacquisition
12	Doppler	Used for reacquisition
13	Prompt Code offset	Used for reacquisition
14	Acquisition finish time	Used for reacquisition

in 0.1 chip steps (a total of 10230 points), at the best estimated Doppler. These two files are useful for sanity-checking the levels of the signals found and tweaking the detection ratio threshold for weaker signals. However, be aware that generating them significantly slows down the acquisition processing.

- Tracking outputs (fastgps_tracking_log.dat): If configured for normal outputs, the following data is written as a single row every 20 ms, to coincide with the navigation data bit rate. When configured in debug mode, the exact same data is written at the higher rate of once every millisecond (which can result in a very large file for long processing times). The output file format consists of the entries in Table 5.3.
- Navigation (normal output) (fastgps_navigation_log.dat): Currently, the position, velocity, and time estimates of the receiver are output at 50 Hz, starting from the point where attempting a PVT estimate is possible (i.e., The receiver clock has been set and ephemeris has been decoded or external aiding data is available). In addition to the PVT estimate calculated by the receiver, raw measurements and satellite PVT information are also provided for parallel observation and testing of the position solution. Table 5.4 lists the format of the navigation log file.
- Navigation (Google Earth output) (fastgps_Google.kml): The Google Earth file is generated in the kml format, which can be directly loaded into the Google Earth mapping environment. This file is not loaded into the post-processing scripts directly. Information on the details of the kml format can be found at the Google Web site [18].

5.5 fastgps Software Receiver Architecture

The fastgps software receiver core modules (i.e., excluding the GUI files) are controlled in the run_fastgps() function in the fastgps.cpp file in the src/libfastgps subdirectory. This is the function called after the “Start” button is pressed on the main display window, or immediately after the receiver is started via the command line. The receiver quickly reads the configuration and runs on its own over the time interval specified, occasionally displaying a message indicating significant state changes or errors. The messages displayed during execution are all stored in the file fastgps_messages.txt for safekeeping. During execution the receiver uses most of the available CPU, and other applications running at the same time will be

Table 5.3 Format of Both the Normal and Debug Tracking Log File

Column	Value	Comment
0	Process time	Seconds
1	Number of channels	Number of channels allocated to tracking satellites.
Repeated for each of the 12 possible tracking channels (number of entries with valid data specified by “Number of channels” above) (chan = 0 through 11)		
5*chan + 2	PRN Number	Satellite being tracked
5*chan + 3	I prompt	Correlation output
5*chan + 4	Q prompt	Correlation output
5*chan + 5	Doppler frequency	hertz
5*chan + 6	spare	

Table 5.4 Format of the Navigation Log File

<i>Column</i>	<i>Value</i>	<i>Comment</i>
0	Receiver time	In GPS seconds
1	Number of valid measurements	Used in position solution
2	Estimated position estimate, X	WGS-84 reference, meters
3	Estimated position estimate, Y	WGS-84 reference, meters
4	Estimated position estimate, Z	WGS-84 reference, meters
5	Estimated receiver clock bias	Meters
6	Estimated position, latitude	WGS-84 reference, radians
7	Estimated position, longitude	WGS-84 reference, radians
8	Estimated position, height	WGS-84 reference, meters
9	GDOP	
10	Estimated velocity estimate, X	WGS-84 reference, m/s
11	Estimated velocity estimate, Y	WGS-84 reference, m/s
12	Estimated velocity estimate, Z	WGS-84 reference, m/s
13	Estimated receiver clock drift rate	m/s
Repeated for each of the 12 possible tracking channels (chan = 0 through 11)		
10*chan + 14	Navigation data state	0 = not used, 1 = have TOW, 2 = Have EPH
10*chan + 15	Pseudorange	Meters
10*chan + 16	Satellite position, X	WGS-84 reference, meters
10*chan + 17	Satellite position, Y	WGS-84 reference, meters
10*chan + 18	Satellite position, Z	WGS-84 reference, meters
10*chan + 19	Satellite clock bias	Seconds
10*chan + 20	Satellite velocity, X	WGS-84 reference, m/s
10*chan + 21	Satellite velocity, Y	WGS-84 reference, m/s
10*chan + 22	Satellite velocity, Z	WGS-84 reference, m/s
10*chan + 23	Satellite clock drift	s/s

sluggish. The processing time is displayed roughly every half a second, and this can be used to estimate how long one's computer will take to process the duration of data specified in the configuration file.

After the configuration file is read and the internal variables are initialized the receiver will enter its main processing loop. This main processing loop operates on small pieces of data, less than 1 ms, and branches to subfunctions to perform the necessary tasks at predefined intervals. The main loop and a breakdown of the high-level processing states are described in Section 5.5.2.

5.5.1 Timing and Clock Management

The receiver was designed to work with different sampling rates and sampling resolutions. Therefore, the main processing loop is not based on a necessarily familiar interval, such as the 1-ms C/A code period. This may seem a bit unintuitive initially, but it makes no difference in how the receiver operates fundamentally.

As implemented in numerous GPS receivers, the fundamental receiver timing reference, used to time-stamp the signal reception time, is established by synchronizing a receiver "clock" using the first sample of data in a given subframe. The GPS time of this subframe boundary is transmitted in the navigation data message and effectively allows the locally maintained, data sample-based, receiver clock to be set relatively accurately with respect to GPS time. However, this internal receiver clock will not be exactly synchronized to GPS time and must be constantly corrected using the output of the position solution. The inexact frequency of the sampled

data (i.e., the real frequency is not *exactly* as claimed on the RF front-end box, and told to the receiver in the configuration file) will cause the signal receive time to contain a bias with respect to GPS time. However, by constantly correcting the receiver time as part of the navigation solution and propagating it forward as new data is read in based on the data sample frequency, it can be maintained close enough to GPS system time to make measurements.

The second half of the timing problem is determining the time of transmission of the satellite signals at a given data sample. We have recorded the exact sample where the satellite subframe was identified. The time at this subframe, as told to us by the navigation data, combined with data bits and 1-ms-code epochs allows us to calculate the signal transmission time at any sample. This is successful due to the very accurate clocks generating the signals on board the satellites.

For example, if we have tracked the signal for 15 navigation data bits since the subframe was received, plus another five code periods into the next data bit, plus the tracking code phase offset, the time since the subframe can be computed as follows; $\text{time_at_subframe} + 15 \times 20 \text{ ms} + 5 \times 1 \text{ ms} + (\text{code phase}/1023) \times 1 \text{ ms}$. The whole of L1 C/A code navigation depends on the signal tracking code being generated at accurate 1-ms intervals. This computed signal transmission time is also the time used to estimate the satellite PVT information used in the least squares receiver estimation.

These two times—the internal receiver time and the transmit time of the signal—are the fundamental parameters needed to make a pseudorange measurement, which is the most basic component of the navigation solution. An illustration of these two time references and how they are connected is shown in Figure 5.4.

5.5.2 Main Processing Loop

The startup and main processing modules are shown in Figure 5.5. These do not represent individual functions but the high-level processing states that the receiver progresses through, ultimately advancing to the navigation state, where the receiver PVT is estimated.

The initialization of the **fastgps** receiver consists primarily of reading the configuration file, initializing internal variables and performing miscellaneous set up tasks. The initialization sequence is only executed once for every run of the receiver.

5.5.3 Acquisition

The acquisition functions were designed to be flexible. The software receiver normally attempts to find signals by scanning over a wide Doppler shift range and using FFTs to search the entire range of code delays. When a signal is found, a fine Doppler search is used to more accurately determine the signal frequency and to hand over to the receiver tracking loops (where an additional coarse and fine frequency pull-in is performed). The receiver will search for individually specified satellites, or all of them as indicated in the configuration file. After satellites have been found, the acquisition log file can then be used to self-aid the receiver the next time it is run using the same data set. An outline of the acquisition processing is shown in Figure 5.6.

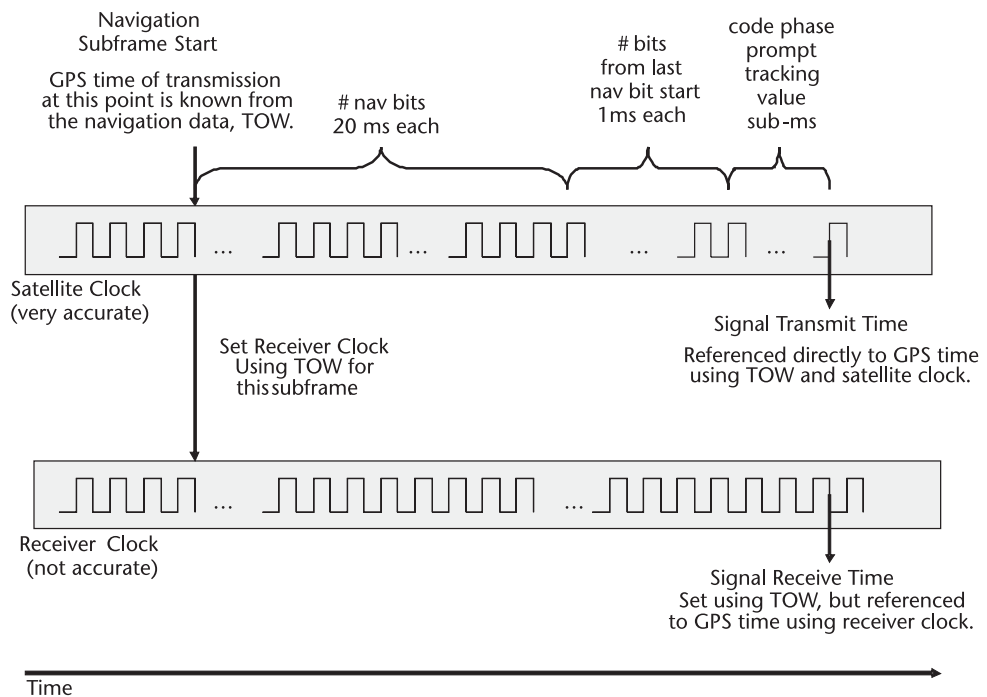


Figure 5.4 Illustration of the two principal time references used in making a satellite pseudorange measurement: the signal transmit time and the signal receive time, both linked to GPS time.

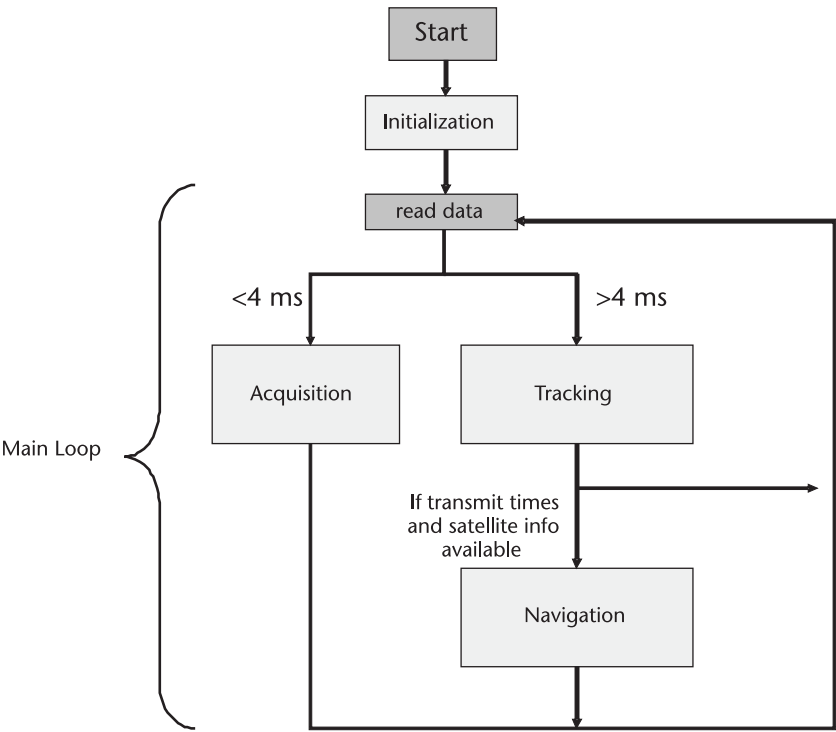


Figure 5.5 Main processing loop within the software receiver, showing functional operating states.

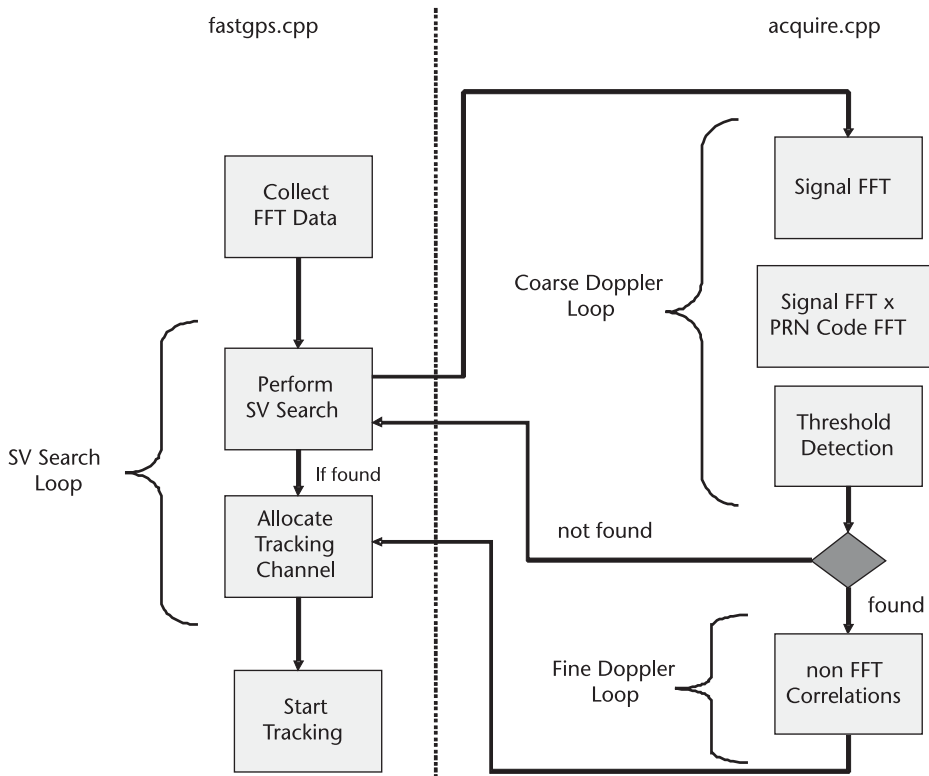


Figure 5.6 The fastgps software receiver acquisition algorithm.

If the receiver has been configured to search for individual satellites, an FFT search technique is used (using the publicly available KissFFT routines [19]) to detect any signals present in the sampled data. More details using FFTs for signal acquisition are discussed in Chapter 2. Alternatively, the receiver could be configured to use the results of the last acquisition file and skip the FFT search, jumping directly to the signal-tracking loops. If not using a self-aiding acquisition file, the receiver executes the following steps.

1. Gather a reasonable amount of data for use in the FFT acquisition processing. The default value is currently 4 ms of data samples (fastgps.cpp).
2. When enough data has been collected, call the acquire function (acquire.cpp), specifying the satellite to search for and scan over a wide range of coarse Doppler bins by performing the following steps.
 - Perform FFT on input sample buffer;
 - Multiply sample FFT and precalculated PRN code FFT;
 - Perform inverse FFT;
 - Search for peaks exceeding the detection threshold.
3. If a signal is found:
 - Perform fine Doppler search-and-store results;
 - Perform debug searches if specified;
 - Allocate it to a tracking channel.

If the receiver finds less than four trackable satellites, it will still proceed to track the signals over the specified duration of data. However, a minimum of four satellites is needed to perform a navigation position estimate. Figure 5.7 shows an example inverse FFT output for a successful satellite search.

5.5.4 Tracking

The tracking functions of the software receiver are divided over several files. At the lowest level we find the software correlator in `correlator.cpp`. This function calls the tracking loops function every 1 ms of sampled data. Subsequently, as the

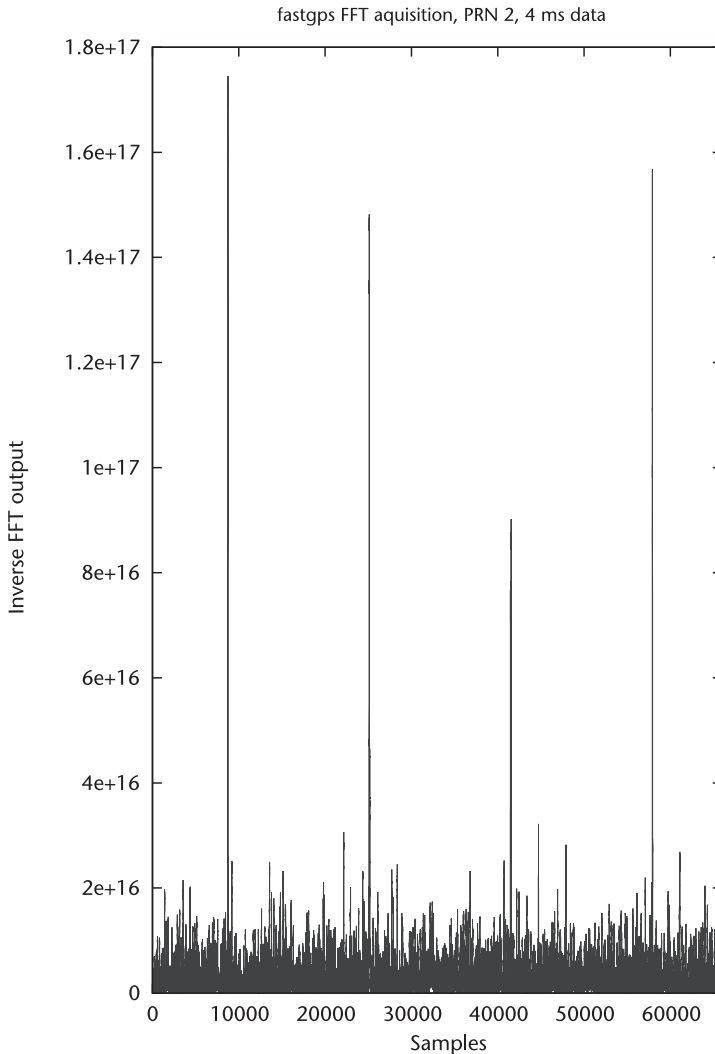


Figure 5.7 Example of the inverse FFT result performed by the **fastgps** software receiver acquisition process. The four peaks represent the GPS PRN code repeating at 1-ms intervals. The largest peak is used to initialize the signal-tracking loops. The four peaks show the signal repeating across the four milliseconds of input data. A data buffer longer than the one millisecond GPS PRN code length is often used to ensure a more reliable signal detection.

signals are tracked and phase lock is achieved, the navigation data subframes can be identified and the data decoded using the routines in `ephemerides.cpp`. Each of these functions is described in more detail below. Figure 5.8 shows an outline of the receiver correlator and tracking loop interaction, performed in parallel for all tracking channels.

5.5.4.1 Software Correlator

The tracking channels are allocated as an array of global data structures, which contain all the variables necessary during processing (called simply “c,” declared in `globals.h` and defined in `structs.h`). The correlator has the task of generating the code and carrier replica signals and performing the necessary correlations and accumulations needed to feed into the tracking functions.

The basic functions of the correlator include, rather simply, incrementally combining the sampled signal with the carrier and code replicas and accumulating the results. When the number of samples processed is equal to a complete code period (1 ms for GPS L1 C/A code) the tracking loop function is called. The correlator continually computes results on six separate correlator spacings for each channel at every sample. These include prompt correlators for the in-phase (I) and quadrature (Q) signals components, as well as early and late offset correlators for

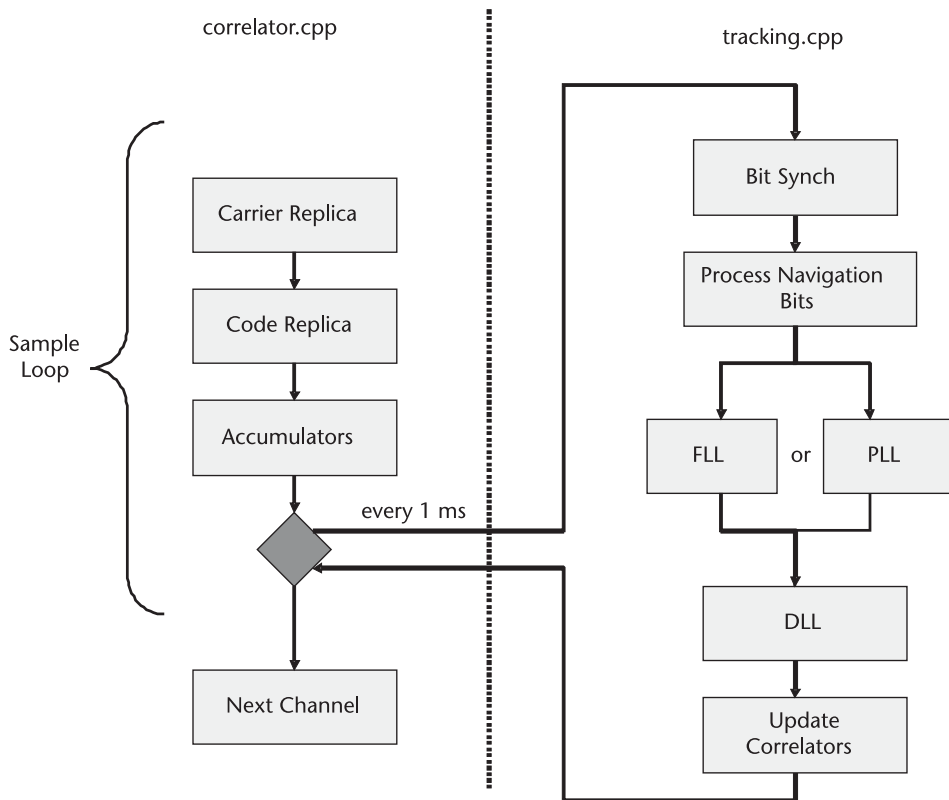


Figure 5.8 The **fastgps** receiver correlator and tracking loops processing.

each. The early and later correlators are spaced one-half chip from the prompt value, and are used primarily in calculating the tracking loop feedback values described next.

5.5.4.2 Tracking Loops

The tracking loops are responsible for performing the initial pull-in of the signals using the information obtained from the acquisition routines, and from there moving to full carrier lock, from which the signal navigation information can be extracted.

At the start of each pass through the loops, the accumulated correlator measurements are used to calculate an error in the replica signal alignment, which is then used to adjust the carrier and code increments used in the correlator for the next set of samples. The tracking error is determined using a discriminator function based on a combination of the six available accumulations calculated for each channel by the software correlator function; see Chapter 2 for more information on discriminator types. This discriminator value is used to calculate new values for the code and carrier increments, which are then fed back into the carrier and code replica generator to be applied during the next correlation interval.

When the tracking loop function is called for a given channel, the tracking time is compared with a number of preset and configurable loop switching times. These times are a frequency locked loop (FLL) switch time, a phase locked loop (PLL) switch time, and a data bit synchronization pull-in time. These parameters determine the switching intervals between the three frequency tracking configurations (FLL high gains, FLL low gains, and PLL). The code tracking delay locked loop (DLL) is performed consistently without any changes over the entire processing interval. This time line is illustrated in Figure 5.9.

During the first interval, a high-bandwidth first-order frequency discriminator is used to quickly correct any coarse errors in the frequency determined by the

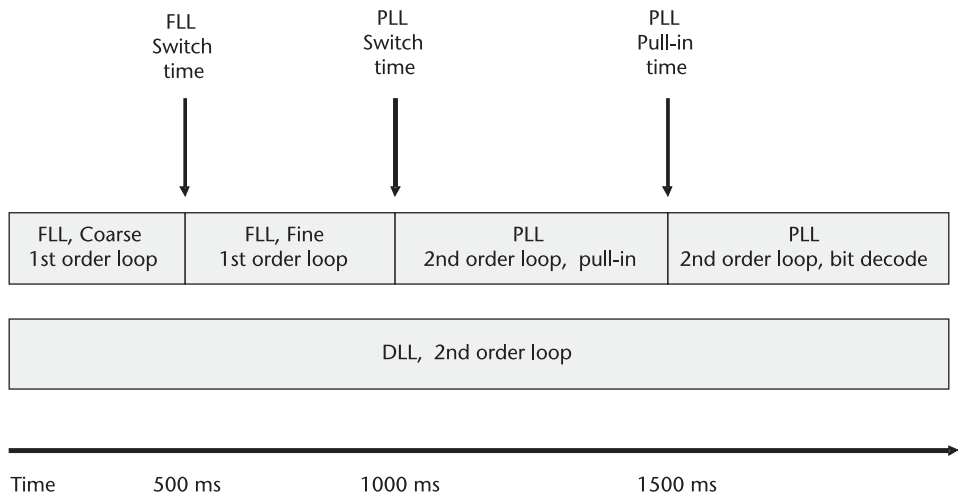


Figure 5.9 The fastgps software receiver-tracking loop timeline.

acquisition functions. The default value is currently set, conservatively, to half a second. As could be observed from the output tracking log files, this is very efficient at quickly correcting large frequency errors, but due to the high gain is very sensitive to noise and will not transition nicely to a PLL. This is resolved by tightening the loop bandwidth (i.e., reducing the gains) after the FLL switch interval, causing the first-order loop to more accurately converge on the actual signal frequency. After another conservative half-second window, we can be confident that the tracked signal frequency is sufficiently close to the real value to transition to the second-order PLL at the PLL switch interval (a total of 1 second into the processing by default). Last, it is usually wise to allow the PLL to settle before processing the in-phase prompt correlation results to determine the navigation data bit edges. This interval is currently set to one-half second, making for a total of 1.5 seconds of processing before the navigation data decoding starts.

It should be noted that the above scheme is very flexible, and many others could easily be substituted and perform equally well or better. However, this method was chosen as it clearly illustrates the subtleties of frequency and phase tracking and how they relate to the frequencies estimated during the acquisition process. The tracking loop gains have been calculated roughly following the guidelines presented by Groves in [12]. For more details on frequency, phase and code delay error discriminators and loop gain estimations, we encourage the reader to consult Chapter 2 or [12]. It is useful to note that because the acquisition performs a relatively tight frequency search, the initial frequency error is usually low when the tracking loops start. However, it would be possible to speed up the acquisition and correct for the additional frequency error using the high-bandwidth FLL. Alternatively, it is possible to accurately determine the frequency during acquisition using an FFT applied over a range of frequencies at the detected code phase delay, this method left as an upgrade for the **fastgps** receiver. An example of a typical frequency tracking sequence is shown in Figure 5.10 for a tracked satellite in data set 1.

Next, after a stable phase lock has been achieved it is then possible to use the in-phase prompt correlator channel accumulations to decode the signal navigation data. However, before we process any navigation data bits, we need to identify accurately the navigation start bit sample. This is accomplished by looking for sign changes in the I prompt signal over 20-ms intervals (the navigation data rate). After the data bit edge has been found, the `process_nav_bit` routine in `ephemerides.cpp` is called every 20m to gather the individual bits and decode them into the navigation message.

5.5.4.3 Navigation Data Decoding

All of the navigation data necessary is decoded in the file `ephemerides.cpp`. This includes only the satellite's time and ephemeris information at the moment, while almanac and other navigation information is ignored. The function `process_nav_bit` acts to manage this decoding, with its first task being to synchronize the incoming bit stream with the known navigation message format [20]. Figure 5.11 illustrates this process.

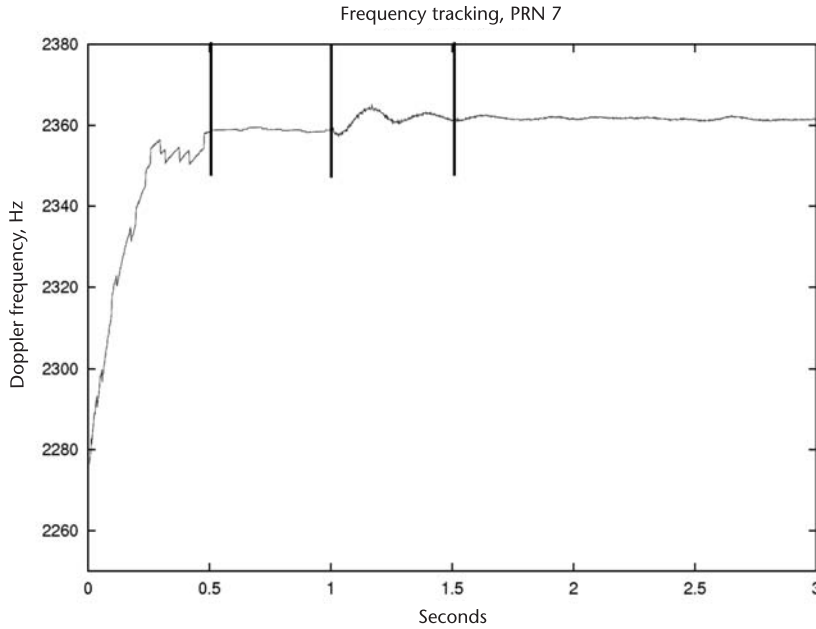


Figure 5.10 The frequency tracking of PRN 7 in data set 1 over the three intervals discussed above. For the first half second the coarse errors in acquisition are corrected. Then, the lower bandwidth FLL brings down the noise in preparation for the switch to the PLL. The last half second (between 1.0 and 1.5 seconds) is the settling period of the PLL, and data bit synchronization interval.

The first task is to find the navigation data preamble patterns. This is an eight-bit sequence (which could be inverted, depending on the phase lock) that marks the start of a navigation data subframe. There are a number of traps lurking in this task, a significant one being that it is often the case that the preamble pattern will occur naturally in the data stream, often in several places that do not mark the start of a subframe. One way to be confident that the preamble one has found is the real thing is to keep processing data until two preambles are found that are separated by exactly the subframe interval (6 seconds for GPS L1 C/A). The software receiver does exactly this by first searching over an entire subframe length and collecting the offsets of all the preambles found. At this point the state machine changes from “SF_SEARCHING” to “SF_VERIFYING.”

After this state is reached, the next preamble we find can be checked to see if it is the real thing by simply looking back 6 seconds and checking if the previous subframe preamble is where it is supposed to be. When a matching pair of subframes has been found exactly 6 seconds apart (and the parity bits have been validated), the state is changed to “SF_RECEIVING,” and the initial subframe can be processed. Note that the first subframe is handled slightly differently than the subsequent ones. When we have found the first subframe we have already decoded the preamble of the next subframe, which puts us 8 data bits ahead of the end of the subframe. By contrast, for the following subframes, we can process the data immediately after the last data bit is received. This slight difference requires some additional logic to be added so we can use the first subframe of data.

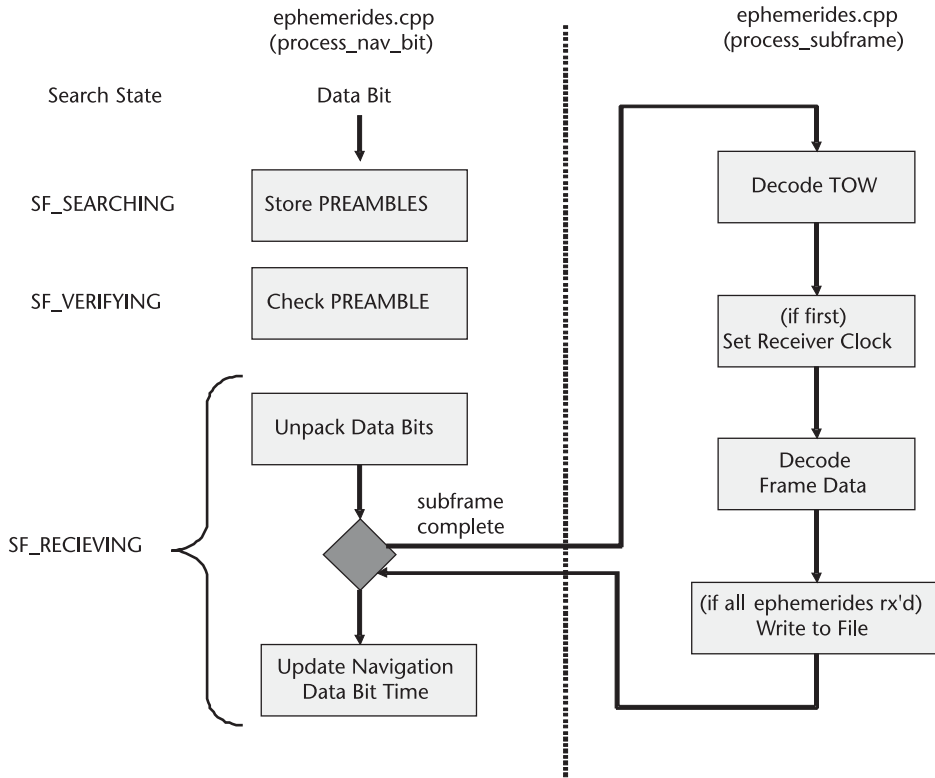


Figure 5.11 The fastgps software receiver ephemeris data-decoding functions.

When a subframe is found, the process subframe routine is called, accomplishing several important things. First, it retrieves the navigation data time of week or TOW of the next subframe start bit (the previous subframe TOW is simply this value minus 6 seconds). This is the key piece of information in determining the received signal time of transmission, discussed in Section 5.5.1. Next, if it has not been done already, this time can be used to set the internal receiver clock, which will be used to make pseudorange measurements. Finally, the rest of the 6 seconds of data is decoded and stored depending on which subframe ID has been received (currently the almanac frames are not decoded). The ephemeris data needed to calculate the transmitting satellite's PVT information is spread out over subframe IDs 1, 2, and 3 and could take over 30 seconds of processing to retrieve all the necessary information [assuming an external aiding file is not available (i.e., from the IGS)]. To speed up future runs with the same data, the ephemeris data for each tracked satellite is saved as it is decoded for reuse. This stored ephemeris can be used on repeat runs with the same data. Alternatively, if an external aiding file is available, the receiver can perform a navigation solution after only one decoded subframe on subsequent runs, for it will only need the signal TOW information to make measurements (note that external time aiding is not currently implemented in the receiver). As the GPS satellite signal transmissions are closely linked in time, the subframes tend to be synchronized across different tracking channels and are often found quite rapidly near the same offset in the data (normally within a few tens of milliseconds).

5.5.5 Navigation

An estimate of the receiver position and velocity can be attempted if the following conditions are met:

- 1. A minimum of four satellites are being tracked.
- 2. The transmission time for each satellite at the measurement data sample is known.
- 3. Ephemeris information for the tracked satellites is available.

When the above conditions have been achieved, the ephemeris is processed and a least-squares navigation solution is attempted. An outline of the various navigation calculation options and how they are managed within the software receiver is shown in Figure 5.12.

The logic shown in Figure 5.12 is an expansion of the navigation block shown in Figure 5.5. The decision-making logic is all executed in the main loop (fastgps.cpp) while all of the navigation-specific support functions can be found in nav.cpp. Alternatively, the control flow on the left calls the functions listed on the right depending on the information available.

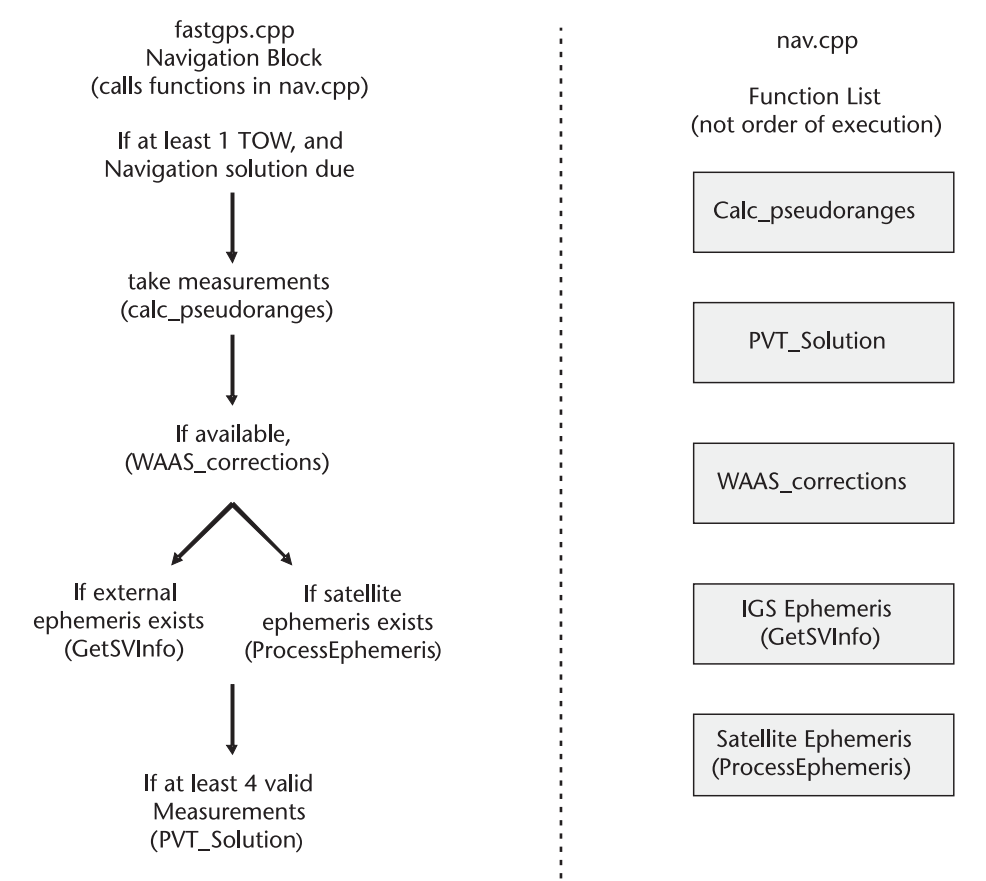


Figure 5.12 The fastgps software receiver navigation solution processing.

5.5.5.1 Pseudorange Measurements

If any subframe has been received on any channel then the receiver clock has been set. This enables the receiver to make a pseudorange measurement for all the tracking channels that have successfully decoded their first subframe, and hence, have a valid transmission time reference. The signal pseudorange is given in (5.1).

$$\begin{aligned} \text{Pseudorange} = & (\text{signal receive time} - \text{signal transmit time}) \\ & \times \text{speed of light} \end{aligned} \quad (5.1)$$

The signal receive time is taken from the internally maintained receiver clock at the measurement sample. Again, any errors in the receiver time will be corrected during the PVT solution as demonstrated in Chapter 3.

The signal transmit time is known with respect to the TOW reference time captured at the first subframe bit. From that point the satellite signal itself is used as a clock to determine the amount of time that has passed from the bit edge reference until the measurement sample. The transmission time is calculated, first by updating the time as each navigation bit is received, and then adding the elapsed milliseconds since the last bit and the submillisecond tcode tracking value, such that,

$$\begin{aligned} \text{time at last bit} = & \text{TOW reference} \\ & + \text{number of bits since TOW reference} \end{aligned} \quad (5.2)$$

$$\begin{aligned} \text{signal transmit time} = & \text{time at last bit} + \text{ms since last bit} \\ & + \text{prompt code}/1023 \end{aligned} \quad (5.3)$$

The signal receive time and signal transmit time calculated here are illustrated in Figure 5.4. These two times, maintained within the receiver for each signal, form the basis for the basic range measurements used during the position estimation. (Please note that to accommodate the early and late correlator spacing, the total code length is offset by a single chip, necessitating a subtraction of one from the code prompt value in the actual implementation.)

5.5.5.2 Differential Corrections

If differential corrections data is available it can be applied directly after the raw pseudorange measurement is made for individual satellites. This can become quite complicated depending on the type of information that is being applied (i.e., atmosphere, satellite clock, and ephemeris, integrity monitoring), but was intentionally kept simple in the **fastgps** receiver. The only differential correction currently implemented is the FAA WAAS real-time ionosphere estimate, using a file of ionosphere grid points (IGP) delays downloaded at approximately the same time as the data was collected. The estimated delay corrections for a given satellite due to the ionosphere at the receiver can be computed using the four corners of the IGP geographic box in which the receiver is located to determine a slant delay correction

(see Chapter 4 for details). These corrections are calculated in the `WAAS_corrections` function in `nav.cpp` using the `IGP_delay.txt` file downloaded at approximately the time of data set 2. IGP delay files can be downloaded from the FAA WAAS web site [15], while a detailed explanation of WAAS is summarized in [16] and documented fully in [17]. The formulas used for the ionosphere corrections in the **fastgps** receiver can be found in Section A.4.4.10.3 of [17].

5.5.5.3 Satellite Information from Ephemeris

Traditionally, in order to perform a PVT solution, the satellite must tell us several things about itself, including its position and velocity as a function of time and the estimated bias and drift rate of its clock. This information is contained in the ephemeris navigation data (subframes 1, 2, and 3). The orbit data is sent as a set of orbital elements, and the clock information as a simple polynomial correction model [20].

As a channel receives its ephemeris data frames the values are decoded and stored. When it comes time to estimate the receiver position, these parameters are passed to the routine `ProcessEphemeris`, which propagates the elements forward or backward, using the widely known method described in [20], and returns the necessary information on the satellites themselves to be used during the PVT solution.

5.5.5.4 Satellite Information from Aiding Data

It is often the case in modern GPS receivers that the ephemeris data as transmitted from the satellite is not available. In other cases, the user may not want to wait up to a half a minute to decode it. This has led to the widespread use of ephemeris aiding or assisted-GPS (AGPS) [21]. This involves gathering satellite ephemeris data (and often timing information) by an external means, then providing it to operational receivers to speed up their time to a full navigation solution, or what is commonly known as the time to first fix (TTFF).

For the case of the **fastgps** receiver, the most convenient way to provide this aiding information is to use the readily available data provided by the International GNSS Service (IGS) [13]. The IGS provides files, applicable for up to 24 hours into the future, containing position, velocity, and clock information for the entire operational GPS constellation. These files can be processed using freely distributed routines (written by Steve Hilla and others at NOAA) to interpolate between the 15-minute data entries. Most IGS data files cover a period of 24 hours (in contrast to the roughly two-hour validity interval for down-linked navigation data ephemeris) and are provided to the GNSS community free of charge. As most PCs running this software receiver will be connected to the Internet, the use of IGS data files has been integrated into the **fastgps** receiver to enable quicker PVT solutions.

For data set 1 described above, the corresponding IGS data file enables positioning after decoding of only a single navigation subframe. For additional data sets collected by users, it is only necessary to note the day on which the data was collected, download the corresponding IGS file, unzip it, and add a line in the receiver configuration file telling the **fastgps** receiver where to find it. Within

the configuration file, the software receiver will then use this IGS sp3 file for the satellite ephemeris information, thus greatly speeding up the time to first fix.

5.5.5.5 Estimating PVT

After all the necessary measurements and satellite information have been gathered together, it is possible to estimate the receiver position, velocity, and clock errors. This is done using the routine `PVT_Solution` in the file `nav.cpp`. This routine is based on the MATLAB/Octave and Python scripts used in Chapter 3, and follows the well-known least-squares estimation procedure described therein. As the raw measurements and satellite information is included in the navigation log file, it is left as an exercise for the reader to reproduce the **fastgps** navigation result using the GNSS simulator processing scripts. See Figures 5.1 and 5.3 for plots of the final navigation position estimates for both included data sets as generated by both the postprocessing scripts and Google Earth.

5.6 Suggested Future Improvements

The following is a short list of some of the more obvious improvements that could be made to future versions of the **fastgps** receiver.

- *Speed improvements.* Several improvements are possible to make the **fastgps** software receiver better live up to its name. A strong contender in this area is to add a code phase measurement only position solution, such as that proposed by Peterson et al [22] and Lannelongue and Pablos [23]. This would allow positioning immediately after signal lock, but requires an additional measurement for the solution of a fifth unknown (coarse GPS time) and places tighter requirements on the a priori position and time knowledge of the receiver. Equally, additional improvements could be made in the efficiency of the correlators and tracking loops.
- *Process additional signals.* Including GPS L2CM and L2CL, GPS L5, as well as Galileo, GLONASS or possibly even Beidou signals. If multiple signals can be processed for a given satellite, it would then be possible to correct for the ionosphere using a combined frequency pseudorange measurement as described in Chapter 3.
- *Improved GUI.* The GUI versions of the receiver provided on the CD for GNU/Linux, Windows, and Mac were intentionally kept very simple. These user interfaces could be enhanced to permit a more interactive run-time environment, possibly including a link to a raw data sampler, from which new data could be captured and processed in near real time.
- *Carrier-smoothing the code measurements.* Carrier-phase tracking is inherently less noisy than the code phase measurements used to make pseudorange measurements. These carrier-phase measurements could be used to smooth the pseudoranges and improve the accuracy of the position estimates.
- *Carrier-phase precise positioning.* A significant amount of work would be required, but it would result in the addition of a very useful feature. In order

to make high accuracy carrier-phase-based pseudorange measurements, the carrier-cycle integer ambiguity would need to be resolved for each signal. These highly accurate measurements would enable additional applications, like those described in Chapter 14 in the fields of geodesy and surveying.

- *Kalman filter positioning.* Many existing receivers rely on Kalman filters to improve the positioning estimation accuracy and robustness. A measurement filter such as this would also facilitate the addition of integrating GPS with external sensors, such as an IMU as described in Chapter 6.

5.7 Further Reading

The following publications provide a small list of other good sources of information on GNSS software receivers (also known as software-defined radios), and we recommend that users looking to explore different implementation options consult these publications.

- *A Software Defined GPS and Galileo Receiver: A Single Frequency Approach* by K. Borre et al., published by Birkhauser 2007. This includes the MATLAB software receiver by D. Plausinaitis et al.
- “Bit-Wise Parallel Algorithms for Efficient Software Correlation Applied to a GPS Software Receiver” by B. Ledvina et al., in the *IEEE Transactions on Wireless Communications*, Vol. 3, Issue 5, September 2004, pp. 1469–1473.
- “Real-Time Software Radio Receiver” by D. M. Akos et al., in *Proc. of ION-GPS National Technical Meeting 2001*.

References

- [1] SourceForge Open Source Software, <http://sourceforge.net/>.
- [2] GPL software licenses, <http://www.gnu.org/licenses/>.
- [3] GNU GCC compiler, <http://gcc.gnu.org/>.
- [4] wxWidgets, cross platform GUI libraries, <http://www.wxwidgets.org/>.
- [5] Bakefile cross platform Makefile generator. <http://www.bakefile.org/index.html>.
- [6] MacOS Xcode IDE, <http://developer.apple.com/tools/xcode/>.
- [7] Python, <http://www.python.org/>, SciPy, <http://www.scipy.org/>, <http://numpy.scipy.org/>, matplotlib, <http://matplotlib.sourceforge.net/>.
- [8] GNU Octave, <http://www.gnu.org/software/octave/>.
- [9] Scilab, <http://www.scilab.org/>.
- [10] R Project, <http://www.r-project.org/>.
- [11] Sparkfun Electronics, GN3S GNSS Sampler V2. http://www.sparkfun.com/commerce/product_info.php?products_id=8238.
- [12] Groves, P., *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*, Norwood, MA: Artech House, 2007.
- [13] International GNSS Service, <http://igsb.jpl.nasa.gov/components/compindex.html>.
- [14] Schenewerk, M. “A Brief Review of Basic GPS Orbit Interpolation Strategies,” *GPS Solutions*, Volume 6, Number 4, 2003, pp. 265–267.

- [15] Federal Aviation Administration Wide Area Augmentation System Web site, <http://www.nstb.tc.faa.gov/>.
- [16] Walter, T., “WAAS MOPS: Practical Examples,” *Proc. of the ION*, National Technical Meeting 1999. http://waas.stanford.edu/mops/MOPS_ex.pdf.
- [17] RTCA document downloads. <http://www.rtca.org/doclist.asp>. RTCA Document Number DO-229A June, 1998.
- [18] GoogleEarth kml file documentation, <http://code.google.com/apis/kml/documentation/>.
- [19] Kiss FFT SourceForge project. <http://sourceforge.net/projects/kissfft/>.
- [20] GPS Interface Control Document, GPS-ICD-200D. <http://www.navcen.uscg.gov/PUBS/gps/icd200/default.htm>.
- [21] Taylor, R. E., and J. W. Sennott, *Navigation System and Method*, U.S. Patent 4445118.
- [22] Peterson, B., R. Hartnett, and G. Ottman, “GPS Receiver Structures for the Urban Canyon,” *Proc. of the ION-GPS*, 1995.
- [23] Lannelongue, S., and P. Pablos, “Fast Acquisition Techniques For GPS Receivers,” *Proc. of the ION-GPS*, 1998.

Integration of GNSS and INS: Part 1

Demoz Gebre-Egziabher, Mark Petovello, and David Bevly

6.1 Introduction

As noted in the introduction and earlier chapters, the purpose of a navigation system is to provide users with an estimate of their navigation state vector. The navigation state vector consists of the user's three position coordinates; the three components of the user's velocity vector, time and, in some instances, the three or more parameters used to describe the user's attitude or orientation. GNSS are designed to efficiently and accurately provide an estimate of the first seven components of the navigation state vector (i.e., position, velocity, and time). GNSS receivers modified to work with multiple antennas can provide the user with an attitude estimate—the final component of the navigation state vector. Thus, GNSS receivers without any additional sensors are capable of providing a complete navigation solution.

The performance of GNSS, however, can be enhanced by integrating them with other navigation sensors or systems. The enhancements provided by such an integration include increased bandwidth (or data rate), robustness, and flexibility and possibly lower overall system cost. Navigation system robustness, for example, can be enhanced by integrating GNSS with some form of dead reckoning. This results in a system that is immune to momentary GNSS outages caused by signal blockage, which is not uncommon in urban environments. Similarly, using a data link to transmit ephemeris information to a GNSS receiver enhances flexibility by allowing acquisition of satellite signals in environments where their power is attenuated to lower than normal levels. This is the A-GNSS approach discussed in Chapter 12. For unmanned aerial vehicle (UAV) operations, low-cost attitude-heading reference systems (AHRS) can be mechanized by fusing GNSS with inexpensive gyros, accelerometers, and magnetometer triads [1, 2].

This and the next three chapters of this book deal with the issues of integration of GNSS with other sensors for the purpose of enhancing overall navigation system performance. This chapter and Chapter 7 consider one popular form of integration, namely the integration of GNSS with inertial sensors. Inertial sensors continuously measure specific force (from which acceleration can be deduced) and rotation rates from which position, velocity, and attitude can be computed. Furthermore, as will be discussed, the error characteristics and vulnerabilities of inertial sensors and GNSS are quite complementary, thus making the two navigation systems ideal for integration.

We begin this chapter with a brief overview of inertial navigation starting from the sensors and continuing through to the modeling of the system errors. Next, the concepts surrounding GNSS and inertial integration are discussed in more detail and the most prevalent integration architectures are presented. The trade-offs associated with the various integration architectures are examined. The estimation algorithms for fusing the GNSS and inertial data are then presented with a particular emphasis on the EKF—the workhorse of navigation sensor fusion algorithms. Practical considerations associated with GNSS and inertial sensor integration are also included. This chapter lays the foundation for the case studies presented in Chapter 7. In Chapter 7, the theory developed here is used to examine various GNSS/INS integration problems.

Before we proceed, it should be noted that it is impossible to provide a complete and in-depth treatment of GNSS-inertial integration in one chapter of a book like this. This and the following chapters are primarily intended to be a quick introduction to the subject with hands-on examples. As such, we encourage the reader to consult other texts on the subject for a more in-depth treatment than is provided here. The literature on the theory behind GNSS-inertial fusion is extensive, and we will closely follow the approach and, to some extent, notation from [3] in this chapter. However, equally excellent pedagogical treatment of the topic can be found in other texts such as [4–11].

6.2 Inertial Navigation

Inertial navigation is a method whereby the current position of a user is determined from a history of acceleration (\mathbf{a}) and angular velocity ($\boldsymbol{\omega}$) measurements. This is accomplished by using the well-known kinematic relationships between acceleration, velocity, and position. That is, acceleration measurements resolved in the appropriate navigation coordinate frame (determined from angular velocity measurements) are integrated once to yield velocity. The computed velocity is then integrated once again to yield position.

6.2.1 Inertial Sensors

Acceleration and angular velocity are measured continually using sensors called accelerometers and gyros, respectively. These two distinct sensors are collectively called inertial sensors. Inertial navigation normally requires orthogonal triads of accelerometers and gyros to measure the full acceleration and angular velocity *vectors*. The term inertial measurement unit (IMU) is used to collectively describe a triad of accelerometers and gyros. The term inertial navigation system (INS) is used to describe the combination of an IMU, the navigation algorithms, and the computer that hosts the algorithms. As such, in the context of this chapter, the terms IMU and INS are not interchangeable, although they are sometimes treated as such in the literature.

The term “accelerometer” is somewhat of a misnomer, however, as these sensors actually measure specific force—nongravitational acceleration. A triad of these is used to measure the specific force vector (\mathbf{f}) from which acceleration is

extracted by a vector addition of \mathbf{g} , the local value of the gravity vector. This requires knowledge of the orientation or attitude of the accelerometers relative to the local \mathbf{g} vector. This orientation or attitude information is obtained by processing the angular velocity measurements obtained from the gyros using the kinematic relationship between attitude description parameters (Euler angles, quaternions, rotation vectors or direction cosine matrices; see [3, 9, 12, 13] for details) and angular velocity. Instead of angular velocity some gyros output incremental rotation or integrated angular rate. This book uses the notation α to denote this quantity.¹ In general, low-cost gyros tend to be rate gyros that output ω while the higher cost (and higher quality) gyros tend to output α . The algorithmic differences associated with mechanizing the inertial navigation equations using ω or α are not relevant to our discussion here and the interested reader should consult [3, 4] for more detail about this subtlety.

6.2.2 Coordinate Frames

We deal with a variety of coordinate frames in inertial navigation algorithms. In the context of this chapter, we consider only four frames. A brief description of these frames follows

- The *inertial frame* is a reference frame that does not accelerate or rotate with respect to the rest of the universe [3]. In navigation applications we use a coordinate frame called the Earth-centered inertial (ECI), which has its origin at Earth's center of mass but does not rotate or accelerate with Earth. As noted in [3] this is not a true inertial frame but is sufficient for most of the terrestrial navigation applications with which we normally deal. The inertial frame is identified with the subscript or superscript i .
- The *body frame* is formed by the axes of the IMU, which, in turn, are assumed to align roughly with the front, right, and down directions of motions of the vehicle being positioned. The body frame is identified with the subscript or superscript b .
- The *Earth frame* originates at the Earth's center of mass and has axes aligned to the mean Greenwich meridian and the Earth's mean spin axis, with the third axis forming an orthogonal frame. More specifically, the first axis points from the center of Earth to the mean Greenwich meridian, and the third axis is collinear with Earth's mean spin axis. The second axis completes the right-handed triad. The Earth frame is identified with the subscript or superscript e .
- The *navigation frame* is a locally defined frame whose axes point in the north, east, and down directions. The navigation frame is identified with the subscript or superscript n .

1. This quantity is normally referred to as “delta theta” and sometimes the symbols $\Delta\theta$ or $\mathbf{\Delta\theta}$ are used to represent it.

6.2.3 Mechanization Equations

The heuristic description of inertial navigation given in the last section is mathematically formalized by the following *equations of motion*, parameterized in the north-east-down (NED) navigation frame

$$\dot{\mathbf{p}} = \mathbf{T}\mathbf{v}^n \quad (6.1)$$

$$\dot{\mathbf{v}}^n = \mathbf{C}_b^n \mathbf{f}^b - [(2\omega_{ie}^n + \omega_{en}^n)^\wedge] \mathbf{v}^n + \mathbf{g}^n \quad (6.2)$$

$$\dot{\mathbf{C}}_b^n = \mathbf{C}_b^n ([(\omega_{ib}^b)^\wedge] - [(\omega_{in}^b)^\wedge]) \quad (6.3)$$

In the above, the rotation rate of frame a relative to frame b , parameterized in frame c is denoted ω_{ba}^c . The vectors \mathbf{p} and \mathbf{v} represent the position and velocity, respectively. The entries of the position vector, \mathbf{p} , are the user's geodetic (or curvilinear) coordinates of latitude (L), longitude (λ) and altitude (h).² Correspondingly, the entries of the velocity vector \mathbf{v} are the NED components of the user's velocity or v_N , v_E , and v_D , respectively. The matrix \mathbf{T} converts the linear velocity values to angular changes in latitude and longitude:

$$\mathbf{T} = \begin{bmatrix} \frac{1}{R_N + h} & 0 & 0 \\ 0 & \frac{1}{(R_E + h) \cos L} & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (6.4)$$

where R_N and R_E are north/south (meridian) and east/west (prime vertical) radii of curvature of the reference ellipsoid [4]. These values are functions of latitude and are computed using standard models. The function *earthrad.m* in the accompanying software calculates these radii as a function of geodetic latitude and altitude.

The local gravity—not gravitation—vector is denoted as \mathbf{g} . The difference between gravitation and gravity is that the latter also includes the effect of centripetal acceleration due to Earth rotation.

A matrix of the form \mathbf{C}_a^b represents the rotation matrix (direction cosine matrix) to rotate a vector from frame a to frame b . Finally, the notation $[(\bullet)^\wedge]$ is the skew symmetric matrix of vector $(\bullet) = [(\bullet)_n \ (\bullet)_e \ (\bullet)_d]^T$ defined as:

$$[(\bullet)^\wedge] = \begin{bmatrix} 0 & -(\bullet)_d & (\bullet)_e \\ (\bullet)_d & 0 & -(\bullet)_n \\ -(\bullet)_e & (\bullet)_n & 0 \end{bmatrix} \quad (6.5)$$

The role of the *mechanization equations* is to perform the numerical integration of the equations of motion to yield the user's updated position, velocity, and attitude

2. We do not use a superscript or subscript on the position vector expressed in geodetic coordinates because there is no ambiguity of coordinate frames. See [3] for a more elaborate discussion of this issue.

parameters. In this regard, the position and velocity parameters are as defined above. However, several options are available for the attitude parameters. For example, if the attitude is described (or parameterized) using Euler angles, the attitude descriptors will be three in number. If quaternions are used [14], they will be four and with a direction cosine matrix they are nine in number. Only three of the attitude states are independent, however, and, thus, if quaternions or direction cosine matrices are chosen to represent the attitude states, constraint equations relating the independent and dependent parameters must be included. In the discussion that follows, we consider an implementation where Euler angles are used. Furthermore, we assume the 3 — 2 — 1 aerospace sequence of yaw (ψ), pitch (θ) and roll (ϕ) for the Euler angles. Thus, the attitude vector, $\boldsymbol{\psi}_{nb}$, is herein given by

$$\boldsymbol{\psi}_{nb} = [\phi \quad \theta \quad \psi]^T \quad (6.6)$$

Note that three Euler angles (which are the ones we will use for the rest of the discussion in this chapter) represent the rotation required to go from the navigation frame to the body frame. Unlike what is done in [3] but for the sake of simplicity, we drop the subscript nb from the Euler angles but will retain it with the attitude vector $\boldsymbol{\psi}_{nb}$. Now that we have described the mechanization equations and associated variables, we can discuss how to implement or solve these equations. Implementing or solving these equations consists of three steps: updating the attitude, updating the velocity, and updating the position, as described below.

6.2.3.1 Attitude Update Equations

The Euler angles used in the mechanization equations describe the orientation of the body frame relative to the navigation frame. Once they are computed, the Euler angles are used to form the rotation matrix, \mathbf{C}_b^n , which allows vectors expressed in the body frame to be mapped or resolved in the navigation frame. This rotation matrix is then used as one of the inputs to the velocity update equations discussed below.

Given a measurement of the angular velocity vector of the body frame relative to the navigation frame expressed (or resolved) in the body frame ($\boldsymbol{\omega}_{nb}^b$), then the Euler angle rates can be computed. That is,

$$\dot{\boldsymbol{\psi}}_{nb} = \mathbf{F}(\boldsymbol{\psi}_{nb}) \boldsymbol{\omega}_{nb}^b \quad (6.7)$$

where

$$\mathbf{F} = \mathbf{F}(\boldsymbol{\psi}_{nb}) = \frac{1}{\cos \theta} \begin{bmatrix} 1 & \sin \phi \sin \theta & \cos \phi \sin \theta \\ 0 & \cos \phi \cos \theta & -\sin \phi \cos \theta \\ 0 & \sin \phi & \cos \phi \end{bmatrix} \quad (6.8)$$

Rate gyros measure the angular velocity of the body frame relative to an inertial frame of reference or $\boldsymbol{\omega}_{ib}^b$. Thus, to determine the attitude of the body frame

relative to the navigation frame (required for attitude determination) we must subtract the rotation rate of the navigation frame relative to the inertial frame, ω_{in}^b , from the output of the rate gyros. That is:

$$\omega_{nb}^b = \omega_{ib}^b - \omega_{in}^b \quad (6.9)$$

For implementations using low-cost gyros, ω_{in}^b is much smaller than the gyro errors that will be discussed in Sections 6.2.5 and 6.5.1. Thus, in this case it is reasonable to assume that $\omega_{nb}^b \approx \omega_{ib}^b$ as measured by the rate gyros. In other words, we assume we are navigating on a nonrotating, flat Earth. For higher quality inertial sensors this assumption is not valid and ω_{in}^b , which is a function of user velocity, must be calculated and subtracted from ω_{ib}^b before computing attitude. This is discussed in more detail in [3, 4].

For applications using automotive or typical consumer grade inertial sensors, a simple Euler integration of (6.7) can be used to generate an estimate of attitude. That is,

$$\begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix}_{k+1} = \begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix}_k + \tau \mathbf{F}(\boldsymbol{\psi}_{nb}(t_k)) \boldsymbol{\omega}_{nb}^b(t_k) \quad (6.10)$$

where $\tau = t_{k+1} - t_k$.³ In some instances, the final product of the GNSS/INS integration algorithm is the attitude solution. In this instance, GNSS/INS is referred to as AHRS. The outputs of AHRS based on low-cost inertial sensors are being used in various applications, most notably for pilot-in-the-loop control of general aviation aircraft [15].

Regardless of the purpose of the integrated GNSS/INS system, the attitude estimate $\boldsymbol{\psi}_{nb}(t_k) = [\phi_k \ \theta_k \ \psi_k]^T$ is essential for computing velocity and subsequently position. More specifically, the elements of $\boldsymbol{\psi}_{nb}(t_k)$ are used to form the transformation matrix \mathbf{C}_b^n as follows:

$$\mathbf{C}_b^n = \begin{bmatrix} \cos \psi_k & -\sin \psi_k & 0 \\ \sin \psi_k & \cos \psi_k & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta_k & 0 & \sin \theta_k \\ 0 & 1 & 0 \\ -\sin \theta_k & 0 & \cos \theta_k \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi_k & -\sin \phi_k \\ 0 & \sin \phi_k & \cos \phi_k \end{bmatrix} \quad (6.11)$$

3. If more sophisticated integration schemes are used, then samples from multiple timesteps are used to propagate the $\boldsymbol{\psi}_{nb}$ forward in time. For example, when a fourth-order Runge-Kutta algorithm is used, then $\boldsymbol{\psi}_{nb}$ at $t = t_{k+1}$ is estimated using inertial sensor samples from t_{k-1} , t_k , and t_{k+1} .

In the software on the DVD that accompanies this book, the m-script *eul2Cbn.m* takes Euler angles as input and computes C_b^n . The matrix C_b^n can now be used in the velocity update equations to compute the user's acceleration vector from specific force measurements made by accelerometers.⁴

6.2.3.2 Velocity Update

The user's velocity is determined by integrating the acceleration (the time rate of change of velocity is acceleration). Depending on what frame derivatives are being taken in, one has to be careful that the *transport theorem* from basic kinematics is applied [16]. Assuming a simple Euler integration is used to update the velocity, then:

$$\mathbf{v}_{k+1}^n = \mathbf{v}_k^n + \tau \dot{\mathbf{v}}^n \quad (6.12)$$

where $\tau = t_{k+1} - t_k$ is the time interval between inertial sensor measurements and $\dot{\mathbf{v}}^n$ is given by:

$$\dot{\mathbf{v}}^n = C_b^n \mathbf{f}^b - [(2\omega_{ie}^n + \omega_{en}^n)^\wedge] \mathbf{v}^n + \mathbf{g}^n \quad (6.13)$$

The first term on the right rotates the specific force measurements from the body frame to the navigation frame. The second term is the correction for Coriolis forces. The Coriolis force accounts for the fact that the NED frame is noninertial. The symbol “ $[(\bullet)^\wedge]$ ” used in the second term is the cross product operator. The variable ω_{ie}^n in the second term is the Earth rate resolved into the local NED frame, and the term ω_{en}^n is the rotation rate of the Earth frame relative to the navigation frame; this latter term is also known as the *transport rate*. Finally, \mathbf{g} is the estimated gravity vector and is usually computed using models for a geodetic reference system such as the international terrestrial reference frame (ITRF) or the World Geodetic System 1984 (WGS84) and in this case is termed the *normal gravity* vector.

At time step $t = t_k$ the Earth rate is computed as discussed in [3]:

$$\omega_{ie}^n \Big|_k = 7.292115 \times 10^{-5} \begin{bmatrix} \cos L_k \\ 0 \\ -\sin L_k \end{bmatrix} \text{ rad/s} \quad (6.14)$$

The function named *earthrate.m* in the accompanying software computes this vector as a function of geodetic latitude. The transport rate is computed as:

$$\omega_{en}^n \Big|_k = \left[\frac{v_E}{R_E + h} \Big|_k \quad -\frac{v_N}{R_N + h} \Big|_k \quad -\frac{v_E \tan L}{R_E + h} \Big|_k \right]^T \quad (6.15)$$

4. We reiterate that the three Euler angles given by $\boldsymbol{\psi}_{nb}$ describe the rotation sequence from NED to body coordinates. However the matrix C_b^n transforms vector from body to NED coordinates. We are being pedantic with this issue because in our experience it has been a source of confusion and many mistakes.

The function *navrate.m* on the accompanying software computes this as a function of \mathbf{v} . The local gravity vector is a function of latitude (L) and altitude (h). Since we are focusing on low-cost inertial sensors here, the following simple model for \mathbf{g} at $t = t_k$ is used [3]:

$$g|_k \approx 9.7803253359 \frac{1 + 1.931853 \times 10^{-3} \sin^2 L_k}{\sqrt{1 - e^2 \sin L_k}} \quad (6.16)$$

where e is the eccentricity of the WGS-84 reference ellipsoid. In the software accompanying this book, (6.16) is implemented in the m-function *glocal.m*. It should be noted, however, that if higher precision is required or higher quality inertial sensors are being used, then more sophisticated models that include corrections for gravity anomalies and deflections of the vertical should be used.

When using low-cost inertial sensors whose measurement errors mask the Coriolis force, (6.13) can generally be simplified to:

$$\mathbf{a} \approx \mathbf{C}_b^n \mathbf{f}^b + \mathbf{g}^n \quad (6.17)$$

As above, this simplification implies that the navigation is being done on a flat, nonrotating Earth.

6.2.3.3 Position Update

The kinematic relationship between position and velocity is given by (6.1). Once again, to first order, a simple Euler integration can be used to propagate the position forward. Note, however, that a precision at the level of meters will require that care be taken in the number of significant digits carried for the values of L and λ (e.g., 1m of north/south displacement is equivalent to approximately $1.6e^{-7}$ rad). With this caveat in mind, the position update equations become:⁵

$$L_{k+1} = L_k + \tau \left(\frac{v_N}{R_N + h} \right) \Big|_k \quad (6.18)$$

$$\lambda_{k+1} = \lambda_k + \tau \left(\frac{v_E}{(R_E + h) \cos L} \right) \Big|_k \quad (6.19)$$

$$h_{k+1} = h_k - \tau(v_D) \Big|_k \quad (6.20)$$

Before continuing, a few remarks are in order. The equations in this section are mostly applicable to low-cost systems where the sensor errors will dominate any errors resulting from approximations made in the mechanization equations. However, for higher quality sensors, some of the simplifications made above may

5. Note that these equations are not valid at the poles where $\cos L = 0$. Also, R_E and R_N are functions of position. Thus, in a more precise formulation, this dependence may have to be taken into account.

not be sufficiently valid and may introduce unmodeled errors into the system. In particular, higher order integrations may be required, and the effects of coning and sculling may also need to be compensated. More information on higher precision integration methods are available in [4, 8–10].

6.2.4 System Initialization

Implicit in the previous section was the assumption that the initial position, velocity, and attitude of the IMU were known. To this end, knowledge of the initial position vector, \mathbf{p}_0 , can be supplied by the GNSS receiver. If the user is static during the initialization period, then the initial velocity, \mathbf{v}_0 , is known and equal to zero. On the other hand, if the initialization is occurring while the user is in motion, then another sensor must be used to supply the initial velocity. Such a sensor may include a GNSS receiver, Doppler radar, air speed sensor (with knowledge of the wind speed), or a wheel speed sensor.

In terms of attitude, if the user is static and relatively high-grade inertial sensors are being used, then the initial attitude, $\psi_{nb}(t=0)$, can be supplied from a leveling (initialization of pitch, θ , and roll, ϕ) and gyrocompassing alignment (initialization of heading or yaw, ψ). However, if low-quality inertial sensors are being used, only leveling is possible. The inability to gyrocompass is due to the fact that the measurement errors of low-cost gyros are too large to sense the Earth's rotation rate [3]. Initialization of ψ in such a system must therefore be accomplished using information from another sensor. This can come from sensors such as a compass or magnetometer triad or, in some cases, by exploiting a known relationship with a velocity sensor (e.g., GNSS) as is done in some automotive applications.

6.2.5 INS Error Model

The equations of motion and corresponding mechanization equations describe the time evolution of the user's position, velocity, and attitude under ideal circumstances. What is generally more interesting from a navigation perspective is to consider the time evolution of the corresponding *errors*. The manner in which errors propagate in an INS can be computed by applying a first-order Taylor series expansion (linearization), or perturbation analysis, to the equations of motion (6.1)–(6.3). Detailed derivations of the equations are widely available in the literature such as [3–7, 10, 11]. For our current purpose, however, we include a slightly simplified form as follows

$$\delta \dot{\mathbf{p}} = \mathbf{T}' \delta \mathbf{p}^n + \mathbf{T} \delta \mathbf{v}^n \quad (6.21)$$

$$\begin{aligned} \delta \dot{\mathbf{v}}^n = & \left[(\mathbf{C}_b^n \mathbf{f}^b)^\wedge \right] \delta \boldsymbol{\psi}_{nb}^n + \mathbf{C}_b^n \delta \mathbf{f}^b - \left[2(\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n)^\wedge \right] \delta \mathbf{v}^n \\ & - \left[(2\delta \boldsymbol{\omega}_{ie}^n + \delta \boldsymbol{\omega}_{en}^n)^\wedge \right] \mathbf{v}^n + \delta \mathbf{g}^n \end{aligned} \quad (6.22)$$

$$\delta \dot{\boldsymbol{\psi}}_{nb}^n \approx - \left[(\boldsymbol{\omega}_{in}^n)^\wedge \right] \delta \boldsymbol{\psi}_{nb}^n + \delta \boldsymbol{\omega}_{in}^n - \mathbf{C}_b^n \delta \boldsymbol{\omega}_{ib}^b \quad (6.23)$$

where a δ indicates the error in the following parameter. In (6.22) and (6.23), the variables $\delta \mathbf{f}^b$ and $\delta \boldsymbol{\omega}_{ib}^b$ are the accelerometer and rate gyro output errors, respectively, which will be discussed in (6.24) and (6.25) shortly. The matrix \mathbf{T}' relates the position errors to their time derivatives. The remaining parameters are as described previously except for $\delta \boldsymbol{\psi}_{nb}^n$. These are not the errors in the Euler angles we defined earlier but attitude errors resolved about the NED frame. The relation of these errors to Euler angle errors will become apparent when we discuss GNSS/INS blending equations later in the chapter in the discussion associated with (6.35). Of note in the above are the errors in the computation of the local gravity vector, \mathbf{g} . This vector is not a constant and varies as a function of location. Position errors lead to errors in \mathbf{g} which, in turn, lead to velocity errors. In particular, errors in the knowledge of the altitude are significant in that they lead to the well-known vertical channel instability problem [7].

The above error equations will be important when discussing the algorithm for fusing the GNSS and INS data (see Section 6.4). For the time being however, the key point is that system performance is ultimately determined by the accelerometer and gyro measurement errors, denoted \mathbf{b}_a^b and \mathbf{b}_g^b respectively. For each sensor, the number of errors is numerous. For simplicity, we assume the measurement can be written as

$$\mathbf{f}^b = \mathbf{f}_t^b + \mathbf{M}_a \mathbf{f}_t^b + \mathbf{b}_a^b + \mathbf{w}_a^b \quad (6.24)$$

$$\boldsymbol{\omega}_{ib}^b = \boldsymbol{\omega}_{ib_t}^b + \mathbf{M}_g \boldsymbol{\omega}_{ib_t}^b + \mathbf{b}_g^b + \mathbf{w}_g \quad (6.25)$$

where $(\bullet)_t$ represents the true value to be measured; $\mathbf{M}_{(\bullet)}$ is a matrix that accounts for the effects of scale factor and nonorthogonality errors; $\mathbf{b}_{(\bullet)}$ is the bias vector; and $\mathbf{w}_{(\bullet)}$ is the uncorrelated output noise. The subscripts a and g represent “accelerometer” and “gyro,” respectively. The superscript b indicates that these errors are expressed in the body frame. In relation to (6.22) and (6.23), note that $\delta \mathbf{f}^b = \mathbf{b}_a^b + \mathbf{w}_a^b$ and $\delta \boldsymbol{\omega}_{ib}^b = \mathbf{b}_g^b + \mathbf{w}_g^b$. Each of these errors (except noise) can be assumed constant, time-varying, or a combination of both (e.g., time variations about a large initial offset). The magnitude and variability of these errors depends on the quality of the sensors, which, in turn, is related to cost. That is, lower cost sensors have generally larger output errors that can vary considerably over a short time period. These output errors limit the performance of inertial navigation. These errors lead to position errors that are a function of time, t , and become larger the longer the inertial navigator operates. Small accelerometer errors will be integrated twice and lead to position errors that grow as a function of t^2 . Angular velocity measurement errors, on the other hand, lead to position errors that grow as a function of t^3 .

In order to mitigate the INS error growth, regular updates from an external measurement source are required. This is the motivation for integrating GNSS with INS. The next section discusses the basic concept of using GNSS for this purpose.

6.3 GNSS/INS Integration Concepts

The motivation of any integrated system is to obtain better performance than would be possible by any of the stand-alone systems. This section begins with a look at the characteristics of GNSS and INS that make them so well-suited to integration. Next, a description of some of the more common integration architectures is given. The mathematical details for fusing the systems together will be examined later in the chapter.

6.3.1 Motivation for GNSS/INS Integration

One of the important advantages of inertial navigators is that they require no interaction with the environment beyond the user. A self-contained, continually operating navigator such as this is desirable, especially to military users where outside aiding cannot be relied upon or is scarce. Contrast this with GNSS, which inherently rely on signals transmitted from satellites. This, of course, cannot be guaranteed in all cases, and temporary outages ranging from seconds to minutes may be possible, depending on the application and operating environment.

In terms of data rate, the rate at which the navigation states can be computed using an inertial navigator is only limited by the processing power available in the INS host computer. Some INS are capable of generating estimates of the navigation state vector at a rate of 100 Hz or more. In contrast, most GNSS receivers have data rates on the order of 1–20 Hz, although some specialized (and more costly) receivers can provide output at up to 100 Hz. Stated differently, the bandwidth of the navigation states generated by inertial navigators is normally much higher than with GNSS, which is indispensable in guidance and control applications and for high-dynamic applications where capturing the full spectrum of motion is important.

GNSS and INS are also complementary in terms of the information they provide. Specifically, although GNSS can be used to provide an attitude solution, this is normally avoided in practice because it involves using a complex, and potentially costly, system with multiple receivers and antennas. In contrast, attitude is a necessary output the INS algorithms.

Finally, and perhaps most importantly, GNSS and INS have complementary error characteristics. The INS output errors are time-correlated and potentially unbounded especially when using low-cost inertial sensors. A GNSS receiver, on the other hand, generates position and velocity estimates with bounded errors. In other words, in terms of their spectra, GNSS contain mostly high-frequency errors, whereas INS errors are larger at low frequency (i.e., the integration performed by the mechanization equations is, effectively, a lowpass filter).

It is also noted that, depending on the GNSS/INS integration architecture used, the INS information can also help increase the robustness of GNSS receiver to jamming or RFI. This involves using INS information to aid the signal processing algorithms inside a GNSS receiver as will be discussed in Section 6.3.2.

Given the above, if the information from both systems are fused, the high-fidelity GNSS position and velocity estimates are used to estimate and arrest the effect of the INS errors. The INS, in turn, provides the high-bandwidth attitude,

position, and velocity estimates needed for vehicle guidance and control. The INS estimates also allow coasting through momentary dropouts of the GNSS solution, which can result from obstructions in the line of sight between the GNSS antennas and the satellites. These represent the primary motivations for integrating the two systems.

6.3.2 Integration Architecture Overview

In general, the integration of GNSS and inertial sensors can be depicted schematically as shown in Figure 6.1. The GNSS receiver either supplies raw signals (down-converted RF samples), raw measurements (observables such as pseudoranges or pseudorange rates), or processed information (position, velocity, and time). An INS (or IMU) supplies the acceleration and angular velocity related information.

The block labeled *integration algorithm* in Figure 6.1 is normally an estimator, which is a mathematical algorithm providing a systematic framework by which information from multiple sources can be fused. Details of these algorithms are deferred to Section 6.4. Rather, this section presents a synopsis of the various GNSS/INS integration architectures, of which there are basically three variants in use today. These are the *loose*, *tight*, and *deep* (also called *ultratight*) integration architectures. While we will stay close to the architecture definitions given in [3, 17], a review of recent literature on this topic will show a lack of consensus regarding the definition of various GNSS/INS integration architectures [18, 19]. This is particularly true for what is called tight integration. Perhaps this is, in part, due to the fact that there are some architectures that contain elements of more than one of the integration schemes such that they cannot be described simply as loose, tight, or deep.

6.3.3 Loose GNSS/INS Integration

Figure 6.2 shows a schematic of the loose GNSS/INS integration architecture. In the loose architecture, the INS and GNSS receivers operate as independent navigation systems. The information from them is blended using an estimator to form a third navigation solution. The information fusion occurs at the PVT level. That is, GNSS PVT estimates are blended with those from an INS.

There are subvariants of this integration architecture. For example it can be designed as a feed-forward or feedback configuration. These are sometimes called

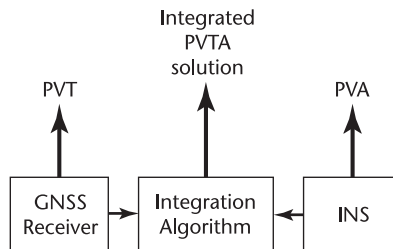


Figure 6.1 GNSS/INS integration (P: position; V: velocity; A: attitude/orientation, and T: time).

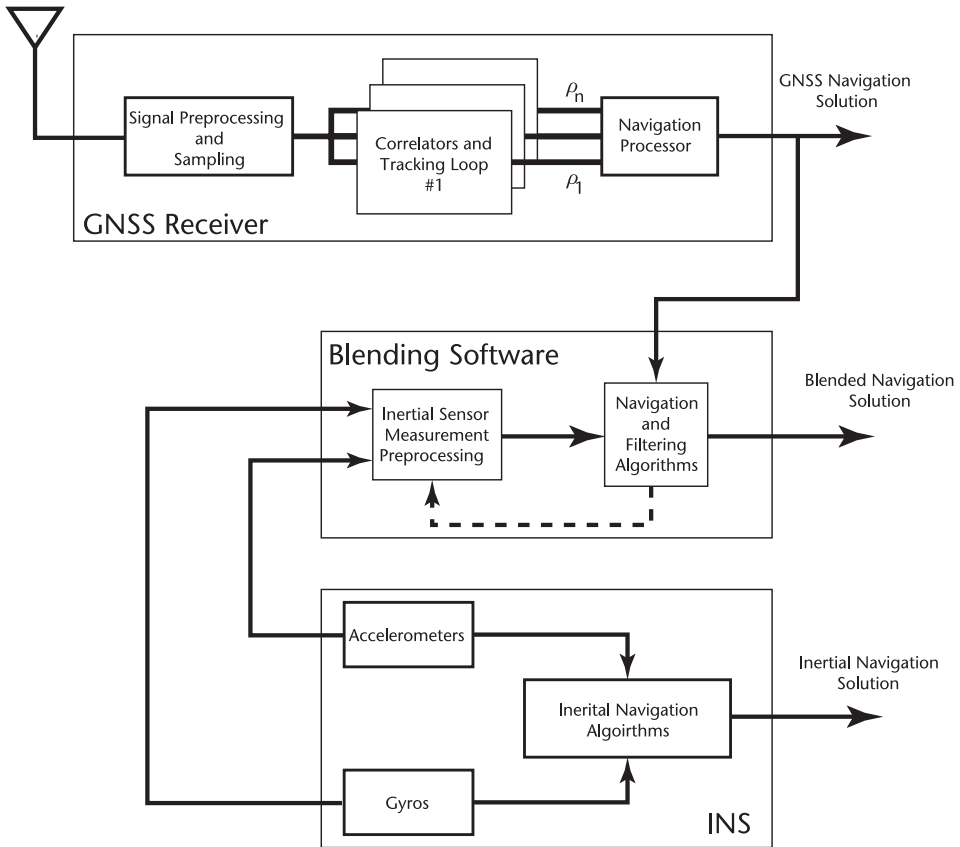


Figure 6.2 GNSS/INS loose integration.

open-or closed-loop configurations, respectively. In the feedback (closed-loop) configuration, inertial sensor errors estimated by the filter are fed back to compensate for the gyro and accelerometer measurement errors. In effect, the sensors are “analytically calibrated” in real-time.

The feedback path is shown by a dashed line in Figure 6.2 to indicate that it is not always required. When it is absent, the system is said to be operating in a feed-forward (open-loop) configuration. The feedback path is almost always required when low-quality inertial sensors (which have large output errors) are used. In this case, if the feedback is absent, the linearization assumption used in the derivation of the INS error model (see Section 6.2.5) may not be sufficiently valid, leading to divergence of the integrated solution. The feed-forward configuration is normally used with high-quality INSs such as those in the navigation grade of sensors. This quality of sensors is what one would normally find, for example, on commercial transport aircraft.

Note that the key feature of the loose integration is that both the INS and GNSS receiver solutions are blended at the PVT level, and the two operate as independent navigators. This is particularly the case when the INS used is of high-quality. Otherwise, the requirement of a feedback path makes the INS dependent on the information from the GNSS receiver. If, for the moment, we focus on loose

integration with a high-quality INS, then the independence noted above and in Figure 6.2 is the key strength of this integration approach. Faults in GNSS do not affect the INS solution (in the short term) and faults in the INS cannot affect the GNSS solution. Another advantage of the loose integration approach is that it is relatively easy to implement because the integration occurs at the PVT level and not the sensor level. As such, very little modification of the basic INS or GNSS sensors or algorithms is needed. This does not mean that no hardware or software work is required. As is always the case with multisensor systems, ensuring that data from both sensors is synchronized or time-tagged correctly can be a very challenging task. Furthermore, the consequences of poorly synchronized measurements can be, at best, an integrated system that performs suboptimally and, at worse, a filter that diverges. More details regarding time synchronization are given in Section 6.6.2.

When low-cost inertial sensors are used in this integration scheme, the INS cannot really be viewed as an independent sensor. Without the feedback that comes from GNSS, the INS solution would diverge very quickly as shown in [20]. In this case the INS brings to the integration attitude information (assuming a single GNSS antenna system is used) and increased bandwidth. The DVD that accompanies this book contains a MATLAB script demonstrating this integration architecture. This script will be discussed in Chapter 7 in the case study, which examines the integration of a low cost IMU with GNSS.

6.3.4 Tight GNSS/INS Integration

A basic schematic of the classic tight GNSS/INS architecture as defined by [3, 21] is shown in Figure 6.3. In this architecture, the INS and GNSS are reduced to their basic sensor functions. That is, pseudorange, pseudorange rate, accelerations, and gyro measurements are used to generate a single blended navigation solution.

In general, the classical tight architecture provides a more accurate and robust solution than loose integration. The enhanced accuracy is, in part, due to the fact that the GNSS information used by the blending filter (i.e., pseudoranges and range rates) exhibit less temporal correlation than the processed PVT solution. It is robust because it can continue to extract useful GNSS information to aid the INSs in situations where less than four satellites are visible. In this instance a stand-alone GNSS solution will not be available but the integrated system will continue to provide one. Of course this will be possible for a limited time only because of GNSS receiver clock drift. If the integration architecture includes additional sensors such as a stable external clock or altimeter, then the duration of useful operation with less than four GNSS satellites can be extended.

In contrast to the loose integration architecture, the main disadvantage of tight integration is that it is more complex and multiple independent navigation solutions do not exist. This latter point is not really a disadvantage particularly when dealing with low-cost inertial sensors. After all, the navigation solution from the low-cost inertial sensors is practically useless without GNSS aiding. Furthermore, this limitation can be dealt with by adding more software to process the GNSS observables and generate a parallel, GNSS-only position, velocity, and time solution.

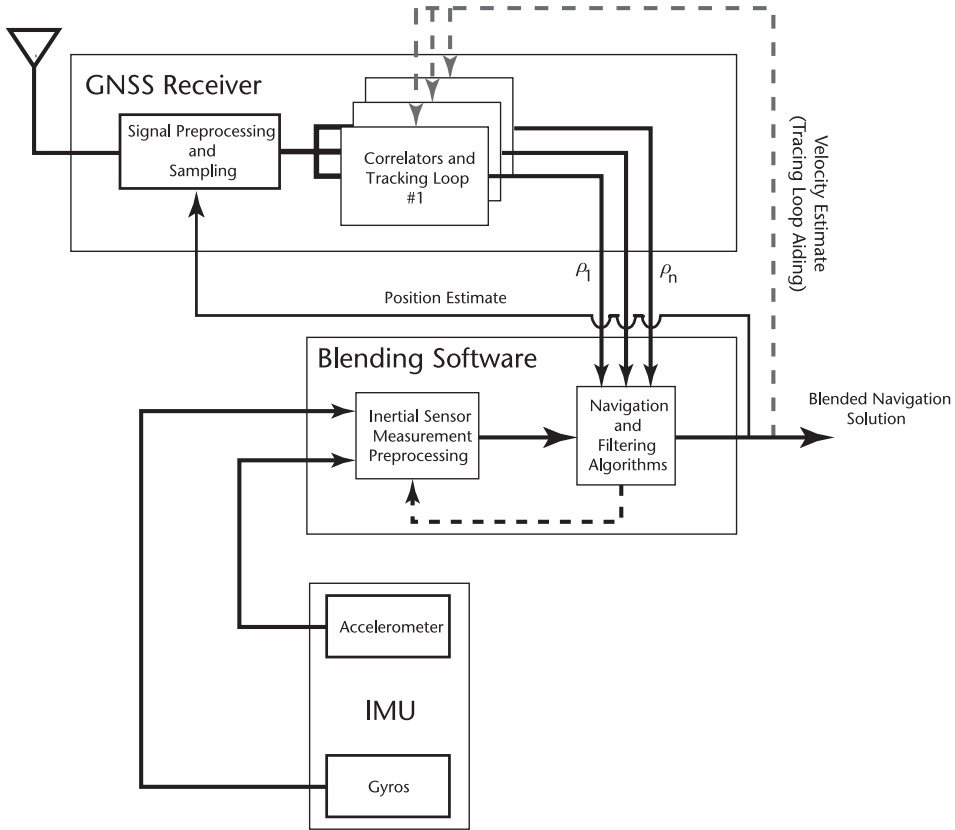


Figure 6.3 GNSS/INS tight integration.

Similar to the loose integration architecture, there are subvariants of the tight integration architecture. The one shown in Figure 6.3 is a variant that is said to feature *tracking loop aiding*. As shown in the schematic, information from the blending filter is fed back to the GNSS receiver to enhance its performance. The information fed back can be position and/or velocity. Position and velocity information from the blending filter can shorten the time required for acquisition or reacquisition of the GNSS signal. On the other hand, velocity (and possibly acceleration) information from the blending filter (shown as a dashed line Figure 6.3) can be used to aid the code and carrier tracking loops in the GNSS receiver. This allows the GNSS receiver to remain in lock (i.e., continue to track the signal) in high-dynamic maneuvers, which would be impossible, or at least difficult, without the aiding information. Similarly, tracking loop aiding can be used to narrow the tracking loop bandwidths of the GNSS receiver thereby increasing its robustness in environments where the GNSS signal carrier-to-noise ratio is reduced.

Once again, the performance enhancements offered by GNSS aiding comes at a price. The GNSS aiding adds complexity as it requires incorporating information from the blending filter into the GNSS tracking loops. Furthermore, in the tight integration architecture the GNSS receiver is no longer independent from the INS; a faulty INS sensor will affect the blended solution, which, in turn, will affect the GNSS receiver's performance.

6.3.5 Deep GNSS/INS Integration

In current usage, the term deep integration refers to the architecture shown in Figure 6.4. There are two variants of this integration architecture—the coherent and noncoherent variants [3]. Figure 6.4 shows the noncoherent variant. There are two important features that distinguish this integration scheme from those discussed above. First, the architecture of the GNSS receiver is fundamentally different from the traditional receiver architecture used in the loose and tight integration schemes. In this case, the traditional receiver architecture consisting of a bank of independent code and carrier-tracking loops is replaced by something akin to a single Vector DLL (VDLL) [22]. The VDLL is enclosed in the dashed box shown in Figure 6.4. Second, the INS is an integral part of the GNSS receiver (note how the mechanization equations are a part of the receiver). That is, the GNSS receiver can no longer be viewed as a navigator independent of the INS. The advantage of the deep integration architecture is that it enhances the robustness of GNSS to interference and jamming. The enhancement afforded by this architecture is better than that provided by tight integration. It also represents an optimal fusion of the information from an INS and a GNSS receiver. However, it should be apparent that the major drawback of the deep integration scheme is the complexity involved in integrating INS information with GNSS information deep inside the receiver. Specifically, much of the necessary modifications would require access to the receiver hardware or to a software-defined receiver such as the one discussed in Chapter 5.

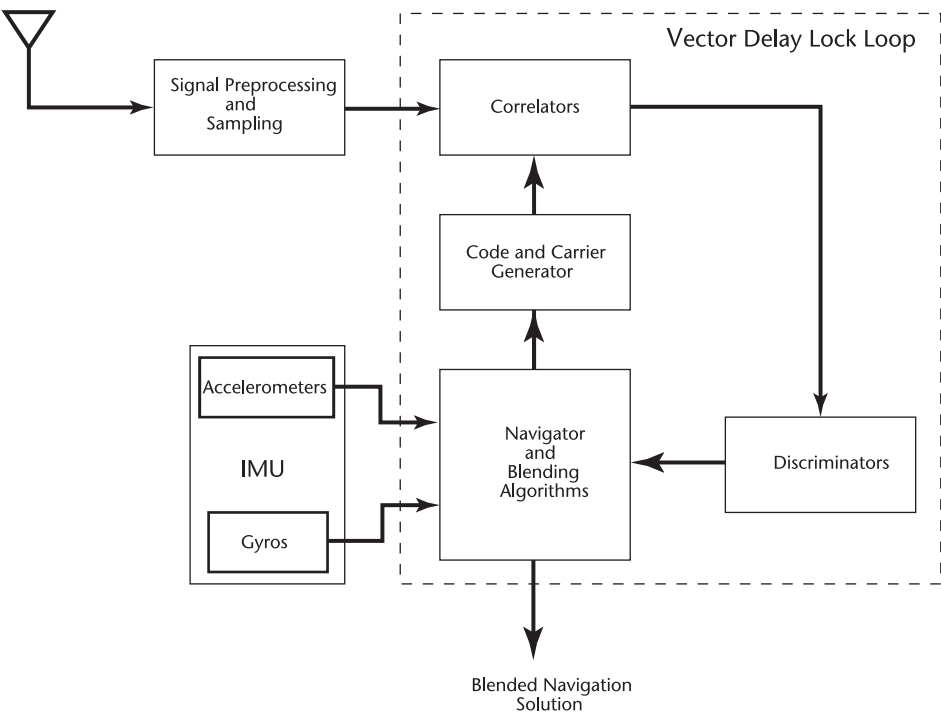


Figure 6.4 GNSS/INS noncoherent deep integration.

In summary, in going from the loose to the deep integration architectures, we gain robustness to GNSS outages either due to vehicle dynamics, interference, or jamming. However, this increased robustness comes at the sacrifice of system simplicity, redundancy and independence of the INS and GNSS navigators.

6.4 Filtering/Estimation Algorithms

All of the integration architectures presented above ultimately require that the INS and GNSS data be fused together. Generally, the block labeled *integration algorithm* in Figure 6.1 is an estimator, or a mathematical algorithm providing a systematic framework by which information from multiple sources can be combined. Perhaps the most widely used estimator for GNSS/INS integration (and the one discussed herein) is the EKF [23–25]. The reasons for this are twofold; Kalman filters generate optimal estimates of the navigation states by minimizing the covariance of the estimation error, and, their recursive form makes them well-suited to efficient implementation. As a matter of fact, in many GNSS receivers, the stand-alone PVT solution is generated using an EKF. In this instance, the velocity information (derived from pseudorange rate, or Doppler shift measurements) is fused with the position information (derived from pseudorange measurements). In contrast to the point solution described in Chapter 3, the EKF represents a better use of information because it leverages the kinematic relationship between position and velocity. It should be noted that other estimation algorithms are also applicable such as the unscented Kalman filter (UKF) and particle filter (PF), which have been considered in some recent applications [26–28], although these are not discussed here.

Section 6.4.1 gives a brief overview of the EKF as it applies to GNSS/INS integration using a closed-loop, loose integration architecture (the interested reader is urged to consult [3, 29, 30] for more details and specifics associated with all of the integration architectures discussed above). The basic equations are presented first, followed by a short discussion about the time evolution of the filter.

6.4.1 Overview of Extended Kalman Filter (EKF) for GNSS/INS

Any type of Kalman filter consists of two parts: a system model and a measurement model. The system model describes the temporal behavior of the states over time, whereas the measurement model describes how observations relate to the states being estimated. The EKF in particular is designed to handle the situation where the system model is nonlinear. From the equations of motion (6.1)–(6.3), the INS states are clearly nonlinear, and linearization is required. Fortunately, the equations for time evolution of INS error equations already provide the linearized model. In other words, the system model for the GNSS/INS EKF is given by (6.21)–(6.23). Handling of the sensor errors (i.e., accelerometer output errors b_a and gyro output errors b_g) is deferred until Section 6.5.1 without loss of generality in the following.

For the time being, however, the state vector, $\delta \mathbf{x}$, can be written as⁶

6. For the tight integration architecture, the loose integration state vector given by (6.26) (denoted explicitly as $\delta \mathbf{x}_{\text{loose}}$), must be augmented by the GNSS receiver clock bias, b_ρ , and clock drift, \dot{b}_ρ , such that:

$$\delta \mathbf{x}_{\text{tight}}^T = \begin{bmatrix} \delta \mathbf{x}_{\text{loose}}^T & b_\rho & \dot{b}_\rho \end{bmatrix}$$

$$\delta \mathbf{x}^T = [(\delta \mathbf{p})^T \quad (\delta \mathbf{v}^n)^T \quad (\delta \boldsymbol{\psi}_{nb})^T \quad (\mathbf{b}_a^b)^T \quad (\mathbf{b}_g^b)^T] \quad (6.26)$$

Having defined the state vector and the system model, we turn now to the *time update*, or propagation, step of the EKF. For an EKF operating in a closed-loop configuration, whenever the state vector is updated, the estimated states (i.e., INS errors) are applied to the corresponding navigation parameters. Correspondingly, the state vector is reset to zero. A consequence of this is that the state vector itself does not need to be propagated forward in time (a null vector multiplied by anything is a null vector). It is still critical however, that the state covariance be propagated using the following equation

$$\mathbf{P}_{k+1} = \boldsymbol{\Phi}_k \mathbf{P}_k \boldsymbol{\Phi}_k^T + \mathbf{Q}_k \quad (6.27)$$

The matrix \mathbf{P}_k is the state error covariance matrix at $t = t_k$ and its elements are the variances and covariances of the individual error states. The matrix $\boldsymbol{\Phi}_k$ is the state error transition matrix and is computed from the INS error equations. The matrix \mathbf{Q}_k is called the process noise covariance matrix and, in the context of INS, represents stochastic information about the sensor errors being estimated. Again, details of this are deferred until later in the chapter.

The EKF measurement or observation model is given by

$$\delta \mathbf{y}_{k+1} = \mathbf{H} \delta \mathbf{x}_{k+1} + \boldsymbol{\epsilon} \quad (6.28)$$

where $\delta \mathbf{y}$ is the measurement error vector, \mathbf{H} is the measurement matrix that relates the observations to the states, and $\boldsymbol{\epsilon}$ are the measurement errors. For loose integration architecture, measurements are formed as the difference between the GNSS and INS position and velocity (i.e., as error estimates). For a tight integration, the measurements will be the difference between the pseudorange measured by the GNSS receiver and the pseudorange predicted based on the INS-estimated position of the user and knowledge of the satellites' locations. For the loose integration case (which is the focus of the hands-on examples on the DVD), the observation matrix becomes

$$\mathbf{H} = [\mathbf{I}_{6 \times 6} \quad \mathbf{0}_{6 \times (M-6)}] \quad (6.29)$$

where $\mathbf{0}_{6 \times (M-6)}$ is a null matrix and M is the total number of states being estimated [e.g., 15 for the state vector given by (6.26)]. Finally, the measurement error vector, $\boldsymbol{\epsilon}$, is characterized by its covariance matrix defined as

$$\mathbf{R} = E\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\} \quad (6.30)$$

where E is the expectation operator. In Kalman filtering theory, the measurement noise, $\boldsymbol{\epsilon}$, is assumed uncorrelated in time. It is a well-known fact, however, that GNSS PVT and measurement errors are time-correlated. This situation can be dealt with in one of two ways. The first, and perhaps the correct way, is to augment the state vector $\delta \mathbf{x}$ with additional states representing the correlated errors. The

second way to deal with this (and the common approach when using low-cost inertial sensors) is to account for the correlation by artificially inflating the measurement noise covariance matrix. This will, of course, will result in a suboptimal solution. However, when low-cost inertial sensors are integrated with GNSS, the objective is not necessarily to get a position and velocity solution that is of a higher quality than the GNSS solution, but rather to increase bandwidth of the position and velocity solution as well as provide attitude. In this case, the tacit assumption is that the GNSS output is “truth.” Furthermore, if the primary objective of fusing GNSS with low-cost inertial sensors is to obtain an attitude solution, then we are relying on the GNSS velocity and position solutions as “truth” anyway. Since without the GNSS-aiding, the attitude solution obtained from the inertial sensors would be useless, and, thus, the assumption that ϵ is uncorrelated is consistent.

The measurement model is used to perform the *measurement update* of the state vector and its covariance matrix when GNSS measurements are available. This update consists of refining our pre-GNSS measurement estimate of the state error $\delta\hat{\mathbf{x}}^-$ (where the “-” indicates a quantity that has not been updated using the most current GNSS data) and state error covariance \mathbf{P}^- . At $t = t_{k+1}$ this proceeds by first computing the gain matrix \mathbf{K}_{k+1} , then updating the state error estimate (i.e., updating $\delta\hat{\mathbf{x}}^-$ to $\delta\hat{\mathbf{x}}^+$) and finally updating the state error covariance matrix (i.e., updating $\hat{\mathbf{P}}^-$ to $\hat{\mathbf{P}}^+$). The equations required to do this are:

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1}^- \mathbf{H}_{k+1}^T (\mathbf{H}_{k+1} \mathbf{P}_{k+1}^- \mathbf{H}_{k+1}^T + \mathbf{R}_{k+1})^{-1} \quad (6.31)$$

$$\delta\mathbf{x}_{k+1}^+ = \mathbf{K}_{k+1} \delta\mathbf{y}_{k+1} \quad (6.32)$$

$$\mathbf{P}_{k+1}^+ = (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_{k+1}) \mathbf{P}_{k+1}^- \quad (6.33)$$

The role of the Kalman gain matrix is to weigh the relative importance of the time-updated states against the information provided by the GNSS measurements. If the GNSS measurements are of a better quality (as determined by comparing \mathbf{P}_k^- with \mathbf{R}_k), then the gain matrix will be large and place more weight on the GNSS solution. Similarly, if the time-updated states are considered more precise, then the Kalman gain is smaller, thus giving less importance to the measurements.⁷

Following the execution of (6.32), the state vector is generally no longer a null vector. For the closed-loop configuration (considered in the hands-on case studies discussed in Chapter 7), the corresponding INS parameters are then corrected for the estimated errors, and the state vector is reset to zero. For all but the attitude parameters, this is accomplished by simply adding the estimated error to the current parameter as follows

$$(\bullet) = (\bullet) - \delta(\bullet) \quad (6.34)$$

7. The gain computation is actually based on the comparison—in the matrix sense—between $\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T$ and \mathbf{R}_k . For the loose integration, the term $\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T$ is equal to the uncertainties in the position and velocity estimates generated by propagating the inertial errors since the last measurement update. In the most general case of Kalman filters, this term represents the state error covariance matrix projected into the measurement space.

For example, the INS position would be updated as $\mathbf{p}_{k+1}^+ = \mathbf{p}_{k+1}^- - \delta\mathbf{p}_{k+1}^+$ where $\delta\mathbf{p}_{k+1}^+$ is a vector of the first three elements of $\delta\hat{\mathbf{x}}_{k+1}^+$.

The attitude errors $\delta\psi_{nb}^n$ cannot be simply added to the previous Euler angle estimates as they are not errors in these angles. Instead, they are corrections to the matrix \mathbf{C}_b^n and must be incorporated using the rotation matrix as follows:

$$\mathbf{C}_b^n(+) = (\mathbf{I}_{3 \times 3} + [\delta\psi_{nb}^n \wedge])\mathbf{C}_b^n(-) \quad (6.35)$$

Finally, the updated Euler angles can be extracted from \mathbf{C}_b^n by:

$$\phi = \tan^{-1} \left(\frac{c_{32}}{c_{33}} \right) \quad (6.36)$$

$$\theta = -\sin c_{31} \quad (6.37)$$

$$\psi = \tan^{-1} \left(\frac{c_{21}}{c_{11}} \right) \quad (6.38)$$

where c_{ij} is the element of \mathbf{C}_b^n in the i th row and j th column.

After the update, $\delta\hat{\mathbf{x}}_{k+1}^+$ is reset to zero and \mathbf{P}_{k+1}^+ is used to initiate another time update cycle.

6.4.2 Time Evolution of a GNSS/INS System

In general, the EKF for the GNSS/INS integration (or even for the stand-alone GNSS PVT solution) behaves in a fashion similar to the depiction in Figure 6.5, which shows the timeline for the overall uncertainty in the estimated states. Initially (at $t = 0$), the navigation parameters are initialized using the best available information and the state error vector is set to zero. Correspondingly, the state covariance

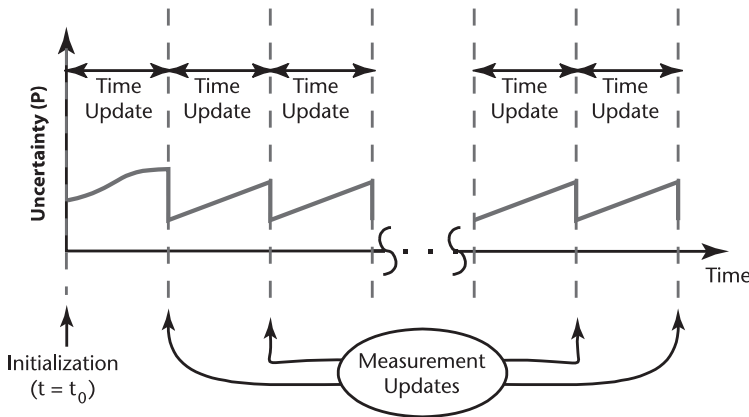


Figure 6.5 Extended Kalman filter event time-line for GNSS/INS integration.

matrix is initialized to reflect the uncertainty in the initial INS parameter estimates. This initial uncertainty is a guess and may be larger than the true uncertainty. That is why the initial uncertainty is shown to be larger in Figure 6.5. Since Figure 6.5 is just a conceptual depiction, the reader should note that it may take several updates (unlike the single update shown in Figure 6.5) to reduce the initial uncertainty to its steady-state level.

During the time update phase, the mechanization equations are used to propagate the position, velocity, and attitude estimates forward in time. Because of the IMU sensor errors, time propagation introduces uncertainty into the system, as reflected by the *upward* sloping lines in Figure 6.5. It follows that the rate of uncertainty increase is directly affected by the quality of the IMU sensors, as will be discussed shortly.

When information from GNSS becomes available, a measurement update is performed. The GNSS data introduces more information to the system and correspondingly decreases the uncertainty in the system. The decrease appears as a *downward* step in Figure 6.5, with the size of the step effectively determined by the Kalman gain matrix. After the GNSS information has been processed to update the state vector and covariance estimates, another time update phase begins.

It is noted that it is not required to alternate time updates and measurement updates. In fact, because the INS solution is typically output at a much higher rate than the GNSS measurements are available, it is common to have several time updates (one per IMU output) between GNSS measurement updates.

6.5 GNSS/INS Integration Implementation

As noted earlier, this book is intended to give the reader a “hands-on” experience with various aspects of GNSS and this chapter is no exception. In what follows, we will use the previous material and walk through the process of implementing a loose GNSS/INS integration filter. We begin by addressing the IMU sensor errors with specific attention given to the selection of errors to be estimated. The section concludes with a brief step-by-step description of how to implement a GNSS/INS system.

6.5.1 IMU Sensor Error Models

Section 6.2.5 presented the INS error model and Section 6.4.1 described how this error model is used in an integrated GNSS/INS system. The handling of the IMU sensor errors however, was postponed until now. As illustrated in (6.24) and (6.25), there are a number of errors affecting the IMU sensors and the question arises as to which of these should be considered significant. In theory, all of the errors are important and should be estimated. Practically, factors such as the desired accuracy, system/processing efficiency, sensor error magnitudes (absolute and relative), level of user dynamics, GNSS update accuracy and frequency, and mission duration will affect a system’s need and ability to reliably estimate all errors. As such, a trade-off is usually made between theoretical veracity on the one hand, and efficient and reliable implementation on the other.

To illustrate this point, let us consider the problem of positioning a vehicle using a low-cost IMU integrated with GNSS. If we consider a typical road vehicle that operates in open-sky environment where GNSS updates are available at regular intervals, say every 1s, then the IMU sensors errors (with the exception of noise) may conceivably be lumped into a single parameter. In this case, the resulting error mismodeling (i.e., treating all the errors together) will be compensated by the regular, high-quality GNSS updates. In contrast, if the vehicle were to operate under foliage where GNSS coverage is sporadic, then such a simplification may be inappropriate, because it does not allow the various errors to be characterized differently, resulting in degraded performance when GNSS is unavailable. Now consider the case where the vehicle is a Formula 1 race car that experiences accelerations of several times gravity (as compared to fractions of gravity in most road vehicles). In this case, the effect of scale factor errors will be considerably magnified and may need to be estimated, even if regular GNSS updates are available. The point to be made here is that there is no hard and fast rule to apply in all situations, and the decision of what states should be estimated must be considered carefully and in light of all available information.

That said, systems based on low-cost inertial sensors generally only consider the bias and scale factor errors for both the accelerometers and gyros. Nonorthogonality errors (and others) are usually excluded because of their small magnitude, especially relative to the bias and scale factor states. Even then, in some cases the scale factors are also ignored, leaving only a single bias term (with the understanding that the “bias” then contains the effect of all other unmodeled errors). We consider this latter case here for illustrative purposes because it is mathematically the simplest, and not necessarily because it is “best.”

It is beyond the scope of this chapter to discuss how to develop mathematical models for the IMU sensors. Instead, we simply cite the results from [31] (using the notation from [3]), which show that for low-cost inertial sensors the accelerometer triad output errors \mathbf{b}_a and rate gyro triad output errors \mathbf{b}_g are a function of time and can be adequately represented by mathematical models of the following form:

$$\mathbf{b}_a(t) = \mathbf{b}_{as} + \mathbf{b}_{ad} + \mathbf{w}_a \quad (6.39)$$

$$\mathbf{b}_g(t) = \mathbf{b}_{gs} + \mathbf{b}_{gd} + \mathbf{w}_g \quad (6.40)$$

$$\frac{d\mathbf{b}_{(\bullet)d}}{dt} = \frac{1}{\tau_{(\bullet)}} (\mathbf{I}_{3 \times 3}) \mathbf{b}_{(\bullet)d} + (\mathbf{I}_{3 \times 3}) \boldsymbol{\mu}_{(\bullet)} \quad (6.41)$$

In (6.41), $\mathbf{I}_{3 \times 3}$ is the 3×3 identity matrix. Also, the symbol (\bullet) is a placeholder for a or g , depending on whether we are dealing with accelerometer or gyro errors, respectively. Note that we have dropped superscripts b from these output errors because it is understood that they are vectors known or expressed in the body frame (where the inertial sensors reside). Some of the components of the output biases are stochastic processes and the notation $\mathbf{b}_a(t)$ and $\mathbf{b}_g(t)$ is meant to indicate that they are time-varying. This should not be interpreted as meaning that these are deterministic functions of time. From these equations it can be seen that the

biases are the sum of three random process. The terms \mathbf{b}_{as} and \mathbf{b}_{gs} represent constant null shifts on the accelerometer and gyro outputs, respectively, which are modeled as random constants. The terms \mathbf{b}_{ad} and \mathbf{b}_{gd} represent a time-varying component of the accelerometer and gyro biases, respectively; sometimes they are referred to as “in-run” biases. They are modeled as an exponentially correlated (first-order Gauss Markov) random process with standard deviation σ_{ad} for the accelerometer errors and σ_{gd} for the gyros.⁸ Thus, while the numerical values for the various parameters will be different for accelerometers and rate gyros, the mathematical form of the error model is the same. Table 6.1 provides a listing of representative values for the bias error model parameters. These values do not reflect any particular sensor but broad classes of sensors as defined by the monikers of “tactical,” or “automotive/consumer,” as defined in [2, 3].

For many applications (particularly when using low-cost inertial sensors), it is normal to estimate the combined effect of null shift and time-varying biases. This fact can be used in constructing a 15 element EKF, the complete 21-state state vector, which can be written as

$$\delta \mathbf{x}^T = [(\delta \mathbf{p}^n)^T \quad (\delta \mathbf{v}^n)^T \quad (\delta \boldsymbol{\psi}_{nb}^n)^T \quad (\mathbf{b}_a)^T \quad (\mathbf{b}_g)^T]^T \quad (6.42)$$

With reference to (6.21)–(6.23), (6.39), and (6.40), the corresponding process noise vector, \mathbf{w}_k , for the integration filter is given by⁹

$$\mathbf{w} = [0_{1 \times 3} \quad (\mathbf{C}_b^n \mathbf{w}_a)^T \quad (\mathbf{C}_b^n \mathbf{w}_g)^T \quad (\mathbf{C}_b^n \boldsymbol{\mu}_a)^T \quad (\mathbf{C}_b^n \boldsymbol{\mu}_g)^T]^T \quad (6.43)$$

Table 6.1 Bias Error Model Parameters for Inertial Sensors

INS Quality	Rate Gyro			Accelerometer		
	σ_{gd} (deg/hr)	τ_g (sec)	σ_n (deg/sec/ $\sqrt{\text{Hz}}$)	σ_{ad} (g)	τ_a (sec)	σ_n (g/ $\sqrt{\text{Hz}}$)
Tactical	0.1–10	100	0.5–0.7	$5\text{--}50 \times 10^{-6}$	60	$50\text{--}100 \times 10^{-5}$
Automotive or consumer	180–360	300	0.1–0.05	$1.2\text{--}2.4 \times 10^{-3}$	100	$0.3\text{--}1 \times 10^{-3}$

8. Note that the time evolution of a scalar random variable η modeled as a first order Gauss-Markov process is governed by the following stochastic differential equations:

$$\dot{\eta}(t) = -\frac{1}{\tau} \eta(t) + \mu(t)$$

The term $\mu(t)$ is a driving white-noise term with a power spectral density S_μ related to the correlation time τ_μ and the variance of the process σ_η^2 by $S_\mu = 2\sigma_\eta^2/\tau_\mu$. Since this is a stochastic differential equation, it cannot be used to determine what η is as a function of time. Instead it is used to determine what the mean and variance of η are as a function of time.

9. If we are dealing with a tight integration architecture, then the process noise vector, $\mathbf{w}_{\text{tight}}$, is nothing more than loose integration process noise vector, denoted here as $\mathbf{w}_{\text{loose}}$, augmented by two more additional noise processes to account for GNSS receiver clock error. That is:

$$\mathbf{w}_{\text{tight}}^T = [\mathbf{w}_{\text{loose}}^T \quad \mathbf{n}_\rho \quad \mathbf{n}_\rho]^T$$

where μ_a and μ_g are white-noise processes with power spectral densities of $2\sigma_{b_{ad}}^2/\tau_a$ and $2\sigma_{b_{gd}}^2/\tau_g$ where τ_a and τ_g are the correlation times for the Markov process error models given in Table 6.1 (see footnote 8). The equation for \mathbf{w} is used to define the process noise power spectral density matrix $\mathbf{Q} = E\{\mathbf{w}\mathbf{w}^T\}$, where E is the expectation operator. In addition to μ_a and μ_g defined above, the diagonal entries of this matrix \mathbf{Q} will consist of the terms $E\{\mathbf{w}_a\mathbf{w}_a^T\}$ and $E\{\mathbf{w}_g\mathbf{w}_g^T\}$, which are the accelerometer and gyro noise power spectral densities given in Table 6.1. The matrix \mathbf{Q} is used to compute the discrete EKF process noise matrix at time step k , or \mathbf{Q}_k using methods described in detail in [3]. With low cost inertial sensor implementations like the ones discussed here, the following approximation can be used:

$$\mathbf{Q}_k = \tau \mathbf{Q} \quad (6.44)$$

where $\tau = t_{k+1} - t_k$. Readers interested in learning more about theoretical details associated with stochastic models are referred to [32–35].

6.5.2 GNSS/INS Integration: Step-by-Step

The first step of any GNSS/INS algorithm is to determine the initial values of position, velocity, and attitude. This is accomplished using whatever data is available at system startup, as discussed in Section 6.2.4. Estimates of the sensor errors can also be initialized using, for example, results of a lab calibration. Alternatively, the sensor errors can be initialized to zero. Equally important is that the covariance matrix of the states also be properly initialized. In general, the diagonal entries of \mathbf{P} are equal to the variances of the various states in $\delta\mathbf{x}$. For the position states, for example, the covariance matrix can be obtained from the estimated accuracy of the GNSS position solution. Similarly, the covariance of the other states is determined based on how their initial values were computed.

Once the system is initialized, the inertial sensors are sampled and the mechanization equations are used to propagate the navigation states forward in time. If operating in a closed-loop configuration, the sensor data should be corrected for the estimated errors before the mechanization equations are executed. The state covariance matrix is propagated forward in step with the mechanization equations using (6.27). If computer throughput is an issue, then the covariance propagation can occur at a slower rate. The IMU data processing is repeated (once for every IMU data sample) until a GNSS measurement is received.

Upon receipt of GNSS measurements (e.g., position and velocity in a loose integration architecture), the error states and covariance matrix are updated using (6.32) and (6.33) and the navigation state vector is updated using (6.34) and (6.35).

6.6 Practical Considerations

Having addressed the fundamentals of GNSS/INS integration, we turn our attention to some of the aspects that must be considered when implementing a practical system.

6.6.1 Lever Arm

Until now, we have implicitly assumed that the IMU and the GNSS antenna are coincident. Generally however, this is not the case and the two are offset in space by some amount. This offset must be accounted for when updating the INS with GNSS information because the two sensors (i.e., the INS and the GNSS receiver) are referring to two physically different points. With this in mind, the lever arm is defined as the rigid vector that runs from the IMU to the phase center of the GNSS antenna and is denoted \mathbf{l}^b . The lever arm is parameterized in the body frame and assumed to be constant over time. The position of the IMU is translated to the GNSS antenna as follows (the opposite translation is performed by moving the second term to the left-hand side)

$$\begin{aligned}\mathbf{p}_{GNSS}^n &= \mathbf{C}_b^n (\mathbf{p}_{IMU}^b + \mathbf{l}^b) \\ &= \mathbf{p}_{IMU}^n + \mathbf{C}_b^n \mathbf{l}^b\end{aligned}\tag{6.45}$$

The accuracy of this translation is a function of the accuracy of the lever arm itself (e.g., how well it is measured and how much it may vary over time) and the accuracy of the attitude solution of the INS. Similarly, the following equation is used for translating the velocity of the IMU to the GNSS antenna

$$\begin{aligned}\mathbf{v}_{GNSS}^n &= \mathbf{v}_{IMU}^n + \mathbf{C}_b^n [(\boldsymbol{\omega}_{nb}^b)^\wedge] \mathbf{l}^b \\ &= \mathbf{v}_{IMU}^n + \mathbf{C}_b^n ([(\boldsymbol{\omega}_{ib}^b)^\wedge] + [(\boldsymbol{\omega}_{in}^b)^\wedge]) \mathbf{l}^b\end{aligned}\tag{6.46}$$

The latter equation reveals that the IMU rotation rates, $\boldsymbol{\omega}_{ib}^b$, are required for the velocity translation. As such, the gyro noise will serve to degrade the quality of the translation, with lower cost sensors introducing more error. Note that filtering of the gyro measurements is theoretically possible but would introduce a time delay, the effect of which would depend on the magnitude of the delay and the level of rotational dynamics.

As already mentioned, the above translations are needed when updating the INS with the GNSS data. In a loose integration, the GNSS position and velocity can be translated to the IMU and the update can be performed as described in Section 6.4.1. In a tight integration however, it is generally simpler to translate the INS position and velocity to the GNSS antenna, instead of trying to project the pseudorange, pseudorange rate, and/or carrier phase data from the antenna to the IMU. In both architectures, the update process is then the same as before.

6.6.2 Timing Requirements

Implicitly assumed until now was that the IMU and GNSS data were time-synchronized. Of course, GNSS measurements are, owing to their inherent time-based derivation, time-tagged. In contrast, IMU measurements are not usually referenced to a specific time base. Details regarding synchronizing the measurements is not discussed here because most systems use dedicated hardware or data collection

systems. Rather, the objective is to provide some insight into what effect a time-synchronization error would have on system performance.

To this end, timing accuracy requirements ultimately relate to the level of dynamics (acceleration) being experienced. To illustrate, consider an IMU timing error of δt relative to the GNSS measurements, which are assumed to be perfectly time-tagged. If the vehicle has constant velocity, both the IMU and GNSS antenna will experience a common position change and thus no error is introduced. However, in the presence of an acceleration, a , the timing error will manifest itself as a position and velocity difference (error) between the two systems/solutions. Assuming a step change in the acceleration for simplicity, fundamental kinematics indicate that the accumulated difference between the GNSS vehicle velocity and the INS-computed velocity is

$$\delta v = a \delta t \quad (6.47)$$

Concordantly, the accumulated position error is

$$\delta x = \frac{1}{2} a \delta t^2 \quad (6.48)$$

Assuming a timing error of 1 ms and an acceleration of 10 m/s², the accumulated position and velocity error would be $\ll 1$ mm and 1 cm/s, respectively. In this case, the position error is negligible regardless of IMU quality. For lower cost systems, the velocity error might also be within the error budget. For higher cost systems however, the velocity error could be problematic because INS-induced errors will be at this level or less (over short time intervals) and the GNSS velocity measurements (in a loose integration system) are also at this level.

In closing, it should be noted that the navigation errors discussed above are not necessarily only present during the actual dynamic period. Specifically, because the EKF interprets this error as originating from one of its states (this is the only information the filter is aware of), it tries to adjust its error estimates accordingly, and it will take time for the error estimates to converge back to their true values.

6.7 Summary and Further Reading

In this chapter we showed that GNSS and INS possess complementary error characteristics and, thus, are well-suited for integration. Various integration architectures were presented. Regardless of the integrated system architecture used, the lower bandwidth GNSS position and velocity information is periodically used to arrest the INS output error drift. These updates also allow calibrating the INS in real time (i.e., on-line estimation of the inertial sensor errors). The INS, in turn, is used to provide the high-bandwidth navigation state information used for vehicle navigation, guidance, and control. In tight and deep integration architectures the INS information is also used to enhance the robustness of the GNSS receiver.

In closing we would like to stress what we said at the outset of this discussion: This chapter is meant to be a guide for the hands-on exercises that accompany

this book. As such, we encourage the reader to exercise the MATLAB-based simulations that accompanies this book. Some of these exercises are discussed in Chapter 7. We would also like to reiterate that this chapter is by no means a complete and thorough treatment of the theory of inertial navigation or Kalman filtering. The reader interested in more detailed treatment of inertial navigation, GNSSs, or filtering should consult the numerous references provided.

References

- [1] Gebre-Egziabher, D., and G. H. Elkaim, "MAV Attitude Determination by Vector Matching," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 44, No. 3, July 2008, pp. 22–33.
- [2] Gebre-Egziabher, D., *Design and Performance Analysis of Low Cost Aided Dead Reckoning Navigator*, Ph.D. dissertation, Department of Aeronautics and Astronautics, Stanford University, Stanford, CA, 2001, p. 56.
- [3] Groves, P. D., *GNSS, Inertial and Multisensor Integrated Navigation Systems*, Norwood, MA: Artech House, 2008.
- [4] Jekeli, C., *Inertial Navigation Systems with Geodetic Applications*, Walter de Gruyter Inc., 2000.
- [5] Farrell, J. A., M. Barth, *The Global Positioning System and Inertial Navigation*, New York, NY: McGraw Hill, 1999.
- [6] Titterton, D. H. and J. L. Weston, *Strapdown Inertial Navigation Technology*, London: Peter Peregrinus, Ltd., 1997.
- [7] Siouris, G., *Aerospace Avionics Systems: A Modern Synthesis*, San Diego, CA: Academic Press, Inc., 1993.
- [8] Savage, P. G., "Strapdown Inertial Navigation Integration Algorithm Design Part 1: Attitude Algorithms," *AIAA Journal of Guidance, Control, and Dynamics*, 21(1), 1998, pp. 19–28.
- [9] Savage, P. G., "Strapdown Inertial Navigation Integration Algorithm Design Part 2: Velocity and Position Algorithms," *AIAA Journal of Guidance, Control, and Dynamics*, 21(2), 1998, pp. 208–221.
- [10] Savage, P. G., *Strapdown Analytics, Volume 1*, Strapdown Associates, Inc., 2000.
- [11] Rogers, M. R., *Applied Mathematics in Integrated Navigation Systems*, 2nd Edition, Reston, VA: American Institute of Aeronautics and Astronautics, 2003.
- [12] Stuelpnagel, J., "On the Parameterization of the Three-Dimensional Rotation Group," *SIAM Review*, Vol. 6, No. 4, 1964, pp. 422–430.
- [13] Hughes, P. C., *Spacecraft Attitude Dynamics*, New York, NY: John Wiley & Sons, Chapter 2, 1986.
- [14] Kuipers, J. B., *Quaternions and Rotation Sequences*, Princeton, New Jersey: Princeton University Press, 2002.
- [15] Crossbow Inc., *NAV 420 GPS-Aided MEMS Inertial System Specification Sheet*, Document Number 6020-0061-01 Rev C.
- [16] Baruh, H., *Analytical Dynamics*, New York, NY: McGraw-Hill, 1999.
- [17] Gebre-Egziabher, D., "What is the Difference between Loose, Tight, Ultra-Tight and Deep Integration Strategies for INS and GNSS?" *Inside GNSS*, Vol. 2, No. 1, 2007. pp. 28–33.
- [18] Pany, T., R. Kaniuth, and B. Eissfeller, "Deep Integration of Navigation Solution and Signal Processing," *Proceedings of the ION-GNSS 18th International Technical Meeting*, Long Beach, CA. September 2005, pp. 1095–1102.

- [19] Pany, T., and B. Eissfeller, "Use of Vector Delay Lock Loop Receiver for GNSS Signal Power Analysis in Bad Signal Conditions," *Proceedings of the IEEE-ION Position, Location and Navigation Symposium (PLANS) 2006*, San Diego, CA, 2006.
- [20] Gebre-Egziabher, D., et al., "An Inexpensive DME-Aided Dead Reckoning Navigator," *Navigation*, Vol. 50, No. 4, 2004, pp. 247–263.
- [21] Greenspan, R. L., "GPS and Inertial Integration," in *Global Positioning System: Theory and Applications*, ed. by Parkinson, Spilker, Enge and Axelrad. AIAA, Washington, D.C., 1996., Vol. 2., pp. 187–220.
- [22] Spilker, J., "Fundamentals of Signal Tracking Theory," in *Global Positioning System: Theory and Applications*, ed. by Parkinson, Spilker, Enge and Axelrad, Washington, D.C.: AIAA, 1996, Vol. 1. pp. 245–328.
- [23] Simon, D., *Optimal State Estimation*, New York, NY: John Wiley and Sons, 2006.
- [24] Stengel, R. F., *Optimal Control and Estimation*, New York, NY: Courier Dover Publications, 1994.
- [25] Bryson, A. E., and Y. C. Ho, *Applied Optimal Control*.
- [26] Shin, E-H, and N. El-Sheimy, "Unscented Kalman Filter and Attitude Errors of Low-Cost Inertial Navigation Systems," *Navigation*, Vol. 54, No. 1, Spring 2007.
- [27] Wendel, J., et al., "A Performance Comparison of Tightly Coupled GPS/INS Navigation Systems based on Extended and Sigma Point Kalman Filters," *Navigation*, Vol. 53, No. 1, Spring 2006.
- [28] Li, Y., et al., "Sigma-Point Kalman Filtering for Tightly Coupled GPS/INS Integration," *Navigation*, Vol. 55, No. 3, Fall 2008.
- [29] Phillips, R., and G. T. Schmidt, "GPS/INS Integration," *AGARD Lecture Series on System Implications and Innovative Application of Satellite Navigation*, Paris, France, July 1996.
- [30] Cox, D. B., "Integration of GPS with Inertial Navigation Systems," reprinted in *Global Positioning System: Papers Published in NAVIGATION*, Institute of Navigation, VA., Vol. I, 1980, pp. 144–153.
- [31] Xing, Z., and D. Gebre-Egziabher, "Modeling and Bounding Low Cost Inertial Sensor Errors," *Proceedings of the IEEE-ION Position Location and Navigation Symposium*, Monterey, CA, May 2008, pp. 1122–1128.
- [32] Gelb, A., *Applied Optimal Estimation*, Cambridge, MA: The MIT Press, 1974.
- [33] Brown, R. G., and P. Y. C. Hwang, *Introduction To Random Signals and Applied Kalman Filtering (Second Edition)*, New York, NY: John Wiley & Sons, Inc., 1992.
- [34] Minkler, G., and J. Minkler, *Theory and Application of Kalman Filtering*, Palm Bay, FL: Magellan Book Company, 1993.
- [35] Grewal, M. S., and A. P. Andrews, *Kalman Filtering Theory and Practice Using MATLAB (Third Edition)*, New York, NY: John Wiley & Sons, Inc., 2008.

Integration of GNSS and INS: Part 2

Demoz Gebre-Egziabher, Mark Petovello, and David Bevly

7.1 Introduction

This chapter continues the examination of integrated GNSS/INS systems and deals with the “hands-on” aspect of GNSS/INS integration. To this end we present three hands-on case studies. The data and MATLAB scripts required to reproduce the results presented and discussed in the chapter come with the software that accompanies this book. The readers are encouraged to follow along using these software examples.

The first case study will be a performance analysis of a low-cost, integrated GNSS/INS navigator designed as part of the navigation, guidance, and control suite for a small UAV. This is followed by a case study motivated by an automotive application. This is an application where an integrated GNSS/INS system is used to estimate vehicle sideslip, which is an important parameter in many electronic stability control (ESC) systems. This case study will include a discussion and software demonstration of issues associated with state observability (particularly heading) in integrated GNSS/INS systems. These two case studies are applications where low-cost inertial sensors of automotive- or consumer-grade quality provide the requisite performance. In closing the chapter, however, we will switch gears slightly and focus on high-accuracy applications, which often use (although do not necessarily require) higher quality (navigation grade) inertial navigators. This final case study deals with the problem of enhancing the integer ambiguity resolution in carrier phase differential GNSS applications.

7.2 Case Study 1: Low-Cost GNSS/INS Integrated Navigator

As noted in Chapter 6, one area where the application of GNSS/INS integration using low-cost inertial sensors has been successful is in the area of guidance, navigation, and control suites for small aerial vehicles. In this application, the purpose of integrating the inertial sensors with GNSS is primarily for providing a means of attitude determination using a single GNSS receiver/antenna. An added benefit of this integration is increased bandwidth of the position and velocity solution than would be available when using GNSS alone. The attitude and higher bandwidth position and velocity solutions are used for various applications

including automatic control of vehicles. This is possible because inertial sensors can easily provide measurements of angular velocity and specific force at rates in excess of 50 Hz.

Figure 7.1 is the ground trajectory of a small aerial vehicle that was equipped with inertial sensors of the automotive-grade quality and a GNSS receiver.¹ The vehicle is a small hand-launched flying-wing described in [2]. Figure 7.2 shows the output of the MATLAB-based GNSS/INS integration algorithm that comes with this book. These plots are a seven minute snippet from the time history for the attitude estimate generated by this system. The figure shows yaw (ψ_{nb}), pitch (θ_{nb}) and roll (ϕ_{nb}) estimates. The attitude histories are shown because with low-cost systems they are the harder states to estimate accurately. From Figure 7.2 we see that the yaw or heading angle is constantly changing. This is consistent with the ground track shown in Figure 7.1 (i.e., a continuous orbit about a point). Figure 7.3 shows the estimates of the accelerometer and rate gyro biases for the same time period shown in Figure 7.2. After the initial transient the estimates settle down to a steady state value. Note that the magnitude of the bias estimates is relatively large. For example, the gyro bias b_{g1} (i.e., output error on the gyro aligned with the first body axis of the aerial vehicle) is approximately 0.6 deg/s for most of the

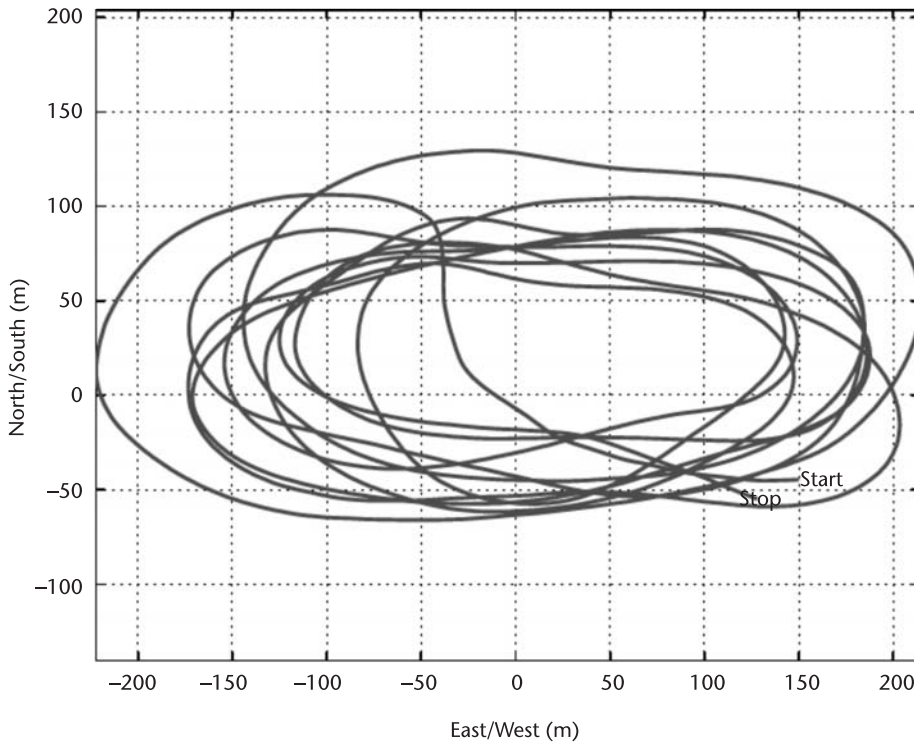


Figure 7.1 Ground trajectory of a small flying wing described in [2].

1. The sensor suite used was the low-cost GPS/INS system known as the Micro Nav produced by Crossbow. At the time of the writing of this chapter, Crossbow announced the discontinuation of the Micro Nav product. However, the performance of the inertial sensors that made up the Micro Nav are typical of many low-cost IMUs. Thus, using another low-cost IMU will lead to more or less similar results.

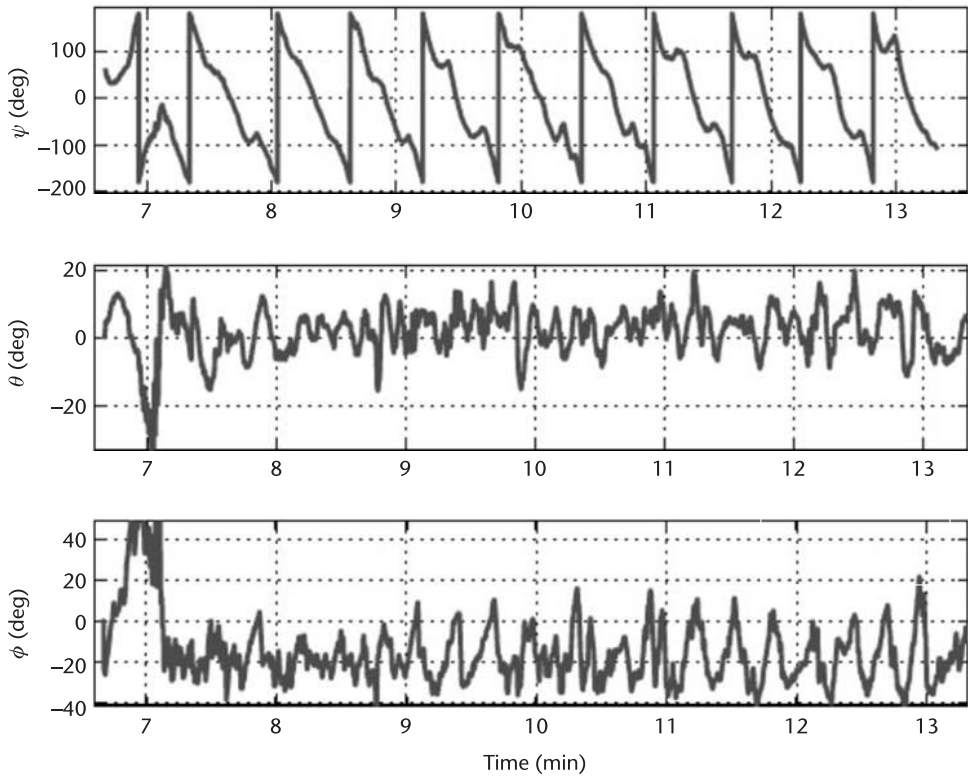


Figure 7.2 Attitude history for a low-cost GNSS/INS integration.

time. This is the combined effect of null shift and time varying biases. Thus, without any aiding, the attitude drift due this rate gyro could be as large as 36 deg/min.

Note the following important point about the bias estimates shown in Figure 7.3. Depending on the initial conditions for the states and the error covariance, and the process noise covariance matrix used, one may get bias estimates that are different that those shown in Figure 7.3. This should not come as a surprise since we are assuming that the sensor errors consist of only additive biases (e.g., we have ignored scale factor errors). Thus, EKF attempts to estimate sensor errors consistent with the model, which may be different from the actual model. While this variability can be reduced by using higher fidelity error models, generating such error models for low-cost inertial sensors is very difficult and not practical.

Returning back to the error estimates of Figure 7.3, we note that a drift rate of 0.6 deg/s is large in the sense that it would lead to a rapid error growth in the navigation state vector estimate without the GNSS aiding. To see the effect of a GNSS outage, we examine the velocity history for the same flight test. Figure 7.4 shows the estimate of the east component of the velocity vector as generated by the GNSS/INS algorithm. We compare this to the GPS generated estimate of v_E^n . The GNSS aiding is turned off at $t = 10$ minutes for a duration of 30 seconds. From Figure 7.4 we see that without GNSS aiding the velocity solution drifts very quickly. The behavior of the GNSS/INS algorithm seen in Figure 7.4 gives credence

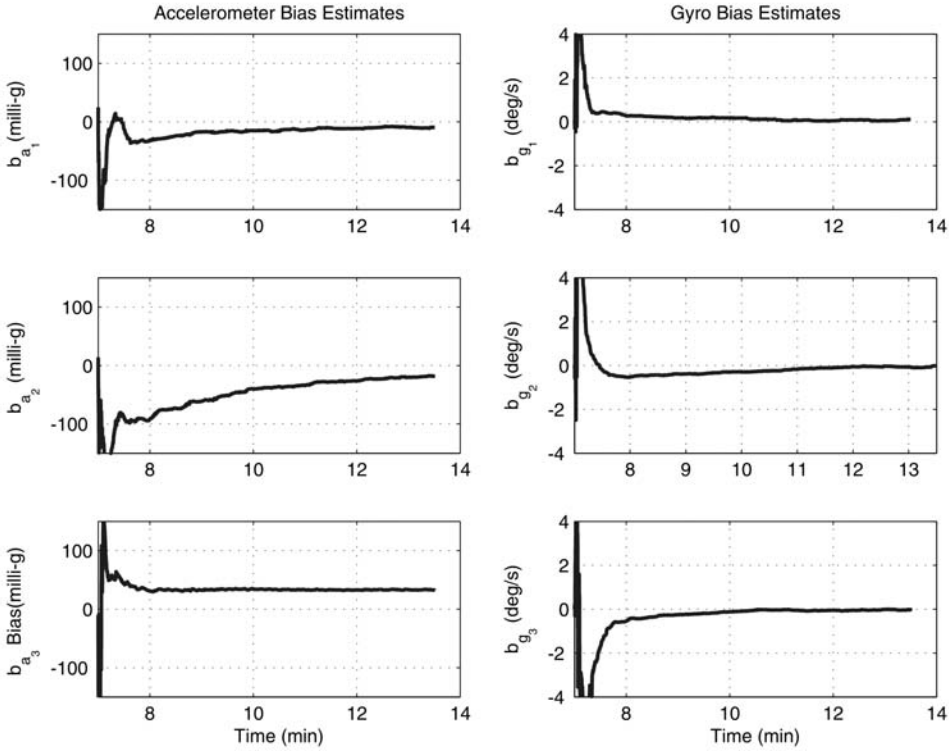


Figure 7.3 Sensor bias estimate for the low-cost GNSS/INS.

to the statement we made in Chapter 6 that the navigation solution from a low-cost inertial sensor suite is practically useless without GNSS aiding.

Note that in Figure 7.4 when GNSS aiding is restored at $t = 10.5$ minutes, the GNSS/INS velocity solution starts converging back to the stand-alone GPS velocity estimate. The rate of convergence is such that it requires approximately 0.1 minutes (or 6 seconds) for the velocity estimates to stabilize. This rate of convergence can be affected by the appropriate “tuning” of the entries in the filter process noise covariance matrices (i.e., changing the values). Care must be taken when doing this, however, as large transients that may result during the convergence period may destabilize the filter and lead to divergence (i.e., estimation errors growing without bound).

Figure 7.5 is a “zoom in” of the GNSS/INS algorithm position estimate. In particular, it shows the estimate of x^e (x -component of the vehicle’s position in ECEF coordinates). It can be seen that the GPS position estimates are periodic and appear once every second (1-Hz update rate). On the other hand, the GNSS/INS algorithm’s output occurs at a rate of 50 Hz. We show Figure 7.5 to highlight that one of the important things an IMU brings to the integrated GNSS/INS solution is higher bandwidth of information in small platforms such as UAVs. It is particularly important that high-bandwidth attitude information be available because it is indispensable for automatic control. Thus, without the integration of GNSS and INSs, it would be very difficult to implement autopilots for small UAVs.

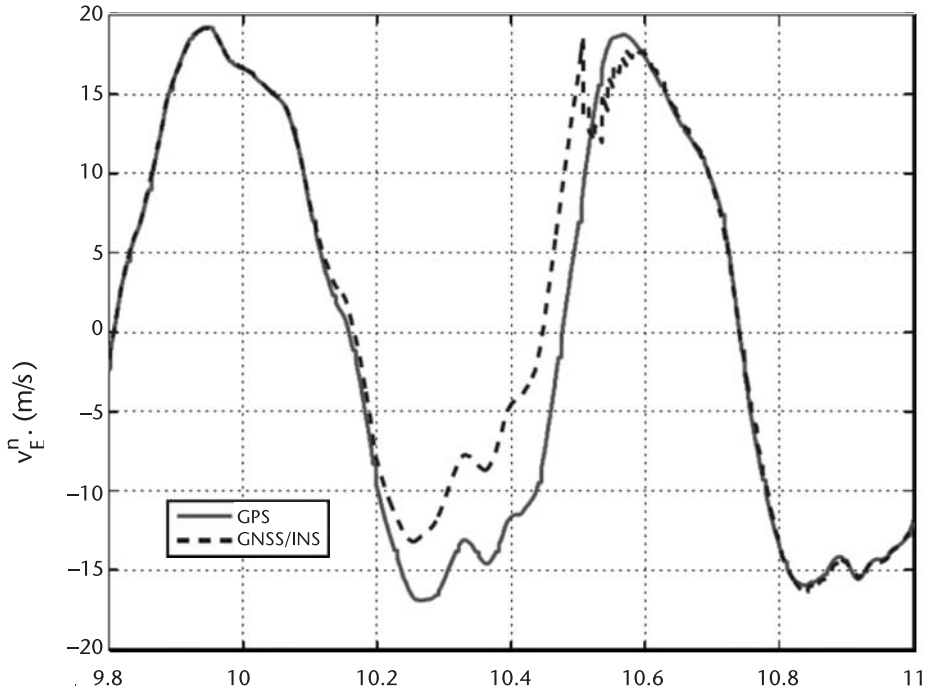


Figure 7.4 Effect of a momentary GNSS outage on the velocity estimate from the GNSS/INS integration algorithm.

7.3 Case Study 2: Vehicle Sideslip Estimation

From the discussions associated with case 1 it is clear that an integrated GNSS/INS algorithm can provide a reasonably accurate, three-axis attitude estimate. Thus, one may conclude that multiantenna GPS attitude determination systems such as the ones described in [3, 4] or ad hoc attitude estimators such as those in [5, 6] are not useful. However, there are instances where the GNSS/INS integration as described in Chapter 6 (or case 1) cannot provide a three-axis attitude estimate. We will examine this aspect of GNSS/INS integration by considering another application where low-cost inertial sensors and GNSS are making a large impact—automotive navigation, guidance, and control. In this regard, this case study examines the motivation for using as well as the limitations of low-cost integrated GNSS/INS systems used to estimate vehicle sideslip as well as critical tire parameters.

7.3.1 Motivation

The sideslip angle (denoted as β here) is the angle between the longitudinal (front-aft) axis of the vehicle and the velocity vector of its center of gravity (CG). Sideslip estimation is a critical aspect of many ESC systems [7–10].

Sideslip estimation schemes fall into two categories, model-based and kinematic. Model based estimators use a dynamic lateral vehicle model such as the bicycle model [11], while kinematic estimators use the kinematic relationships between

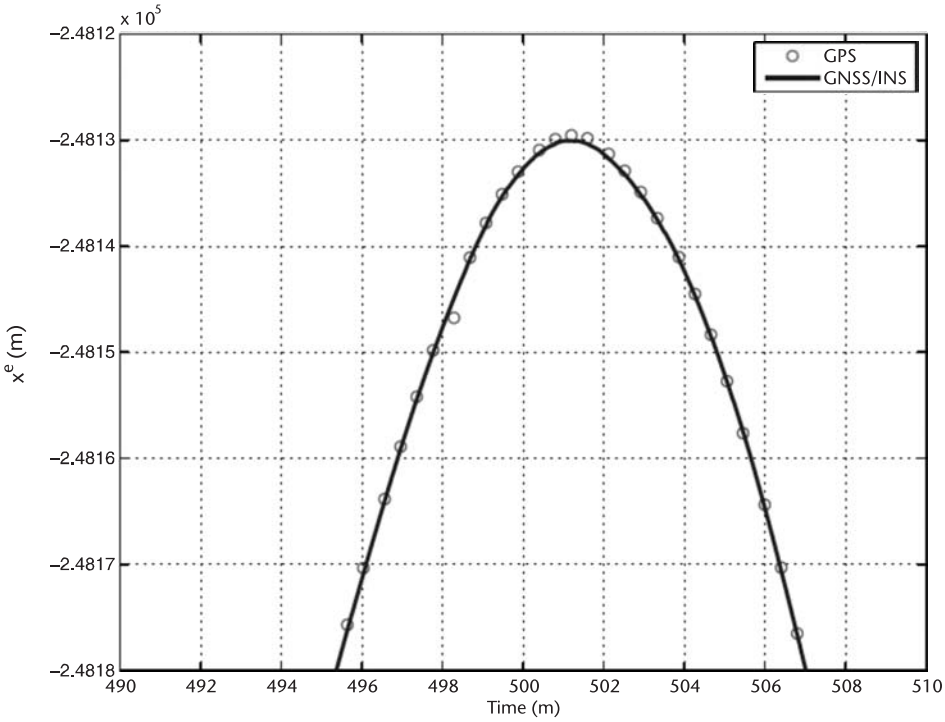


Figure 7.5 Close-up of the ECEF (or Earth frame) x coordinate estimate.

the measurements of different sensors such as GPS and IMUs [12, 13]. Model-based estimators are typically cleaner than kinematic estimators. However, they require knowledge about the vehicle parameters that is often difficult to obtain (such as cornering stiffness). The only vehicle-dependent knowledge kinematic estimators require is the location of the sensors and CG to translate velocities and accelerations to a common point using something akin to (6.45) and (6.46) in the previous chapter. However, even this is often neglected.

A typical method to measure sideslip on test vehicles is using an optical sensor to measure the lateral and longitudinal velocity. However, these sensors are expensive and their performance can be dependent on the surface on which the vehicle is traveling. For these reasons, on production vehicles, this state is not measured. A newer and inexpensive (relative to the optical sensors) way to measure sideslip is using a low-cost integrated navigation system such as the GNSS/INS systems discussed earlier.

The sideslip angle can be estimated as the difference between the vehicle heading and direction of travel (course) measured at the CG. That is:

$$\beta = \tan^{-1} \left(\frac{v_E^n}{v_N^n} \right) - \psi_{nb} \quad (7.1)$$

The first term on the right-hand side of (7.1) can be computed using the outputs of a GNSS receiver or the velocity estimates generated by an integrated GNSS/INS

navigator. The second term, ψ_{nb} , is part of the attitude estimate generated by a GNSS/INS navigator. Thus, as long as attitude and velocity estimates are available, sideslip can be estimated. Figure 7.6 shows the outputs of a sideslip estimator. The figure on the left is the sideslip and heading histories for a vehicle traveling on a low-friction surface (“low μ ”) and the one on the right is for asphalt.

These estimators are Kalman filters, which fuse the information from a vehicle dynamic model with GNSS and INS. Details on these estimators can be found in [14].

The ultimate purpose of sideslip estimation is to predict the cornering stiffness and peak lateral force of the front and rear axles by estimating the lateral force and tire slip angle over the entire operating range. For example, the lateral force estimates generated for an Infiniti G-35 Sedan using sideslip information are shown in Figure 7.7.

Unfortunately, high-quality estimates of β are not always available. This occurs when GNSS measurements are not available, the observability of the ψ_{nb} estimate is poor, or both. The former occurs, for example, in urban canyons where the

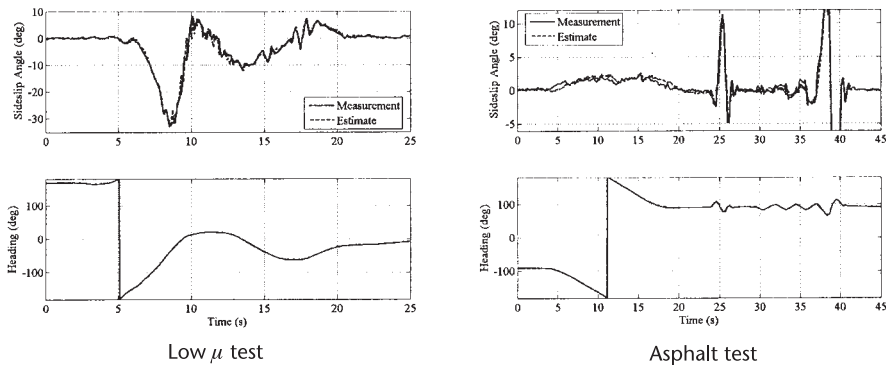


Figure 7.6 Sideslip estimates using a filter-blending GNSS, INS, and a vehicle dynamic model.

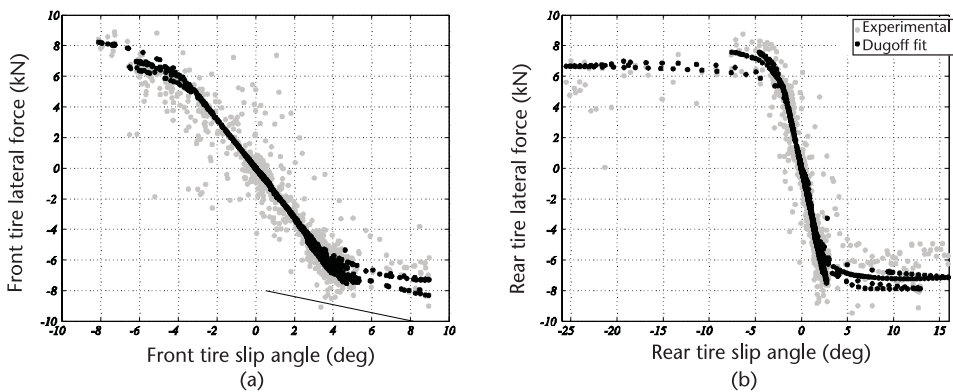


Figure 7.7 (a, b) Lateral force estimates for an Infiniti G-35 sedan [14]. The output of an INS/GNSS integrated navigator is an input to the algorithm, which estimates these side forces.

GNSS signals are weak or obstructed. The latter occurs when the GNSS information is not sufficient to keep errors stemming from inertial sensor biases in check. This latter case is related to the issue of observability, which we will discuss in some detail in Section 7.3.2.

7.3.2 Observability

Information from GNSS may not be able to keep inertial sensor errors in check when one of the following situations arise:

1. Information from GNSS does not have a “path” to get to the inertial sensors and calibrate them. This is a condition known as lack of *geometric* or *deterministic* observability [15–21].
2. Information from the GNSS has a “path” to get to the inertial sensor but cannot arrest the error growth rate of the inertial solution. This is said to be a lack of stochastic observability [16, 18].

In what follows, we will make clear what is meant by “path from GNSS to inertial sensors.” The important point to take away from this discussion is that there are conditions under which the integration of GNSS and inertial sensors will not result in a drift-free attitude solution. This point can be made clearer by considering the specific example of sideslip estimation under the conditions shown in Figure 7.8.

Figure 7.8 shows two different trajectories that are traveled by an automobile at constant speed. In both cases, the vehicles are traveling north, and we invoke the flat, nonrotating Earth assumptions. That is, the vehicle is moving perfectly level (i.e., $h = \text{constant}$ and $\dot{v}_D = 0$) and we are considering short propagation times and motion over short distances. The short propagation times enable us to ignore the complexities associated with a curved Earth (which comes through the matrix \mathbf{T}) as well as ensuring that we do not fall off the hypothetical flat Earth. Since we are assuming a flat, nonrotating Earth and constant speeds, during the constant heading trajectory A ($\psi = 0^\circ$) the vehicle will not experience acceleration. Because of the continually changing heading along trajectory B, however, the vehicle will experience an acceleration. We will consider the case where the inertial sensors used are of the automotive- or consumer-grade quality. Given these approximations, as we will show next, for trajectory A, the heading errors resulting from gyro drift cannot be arrested by GNSS position and velocity measurements while, on trajectory B, they can. This is because on trajectory A, heading lacks observability. To see why heading is unobservable on trajectory A and not trajectory B, let us examine the velocity differential equation. Since the gyro output biases are large and we are assuming a flat Earth, then (6.2) simplifies to:

$$\dot{\mathbf{v}} \approx \mathbf{C}_b^n \mathbf{f}^b + \mathbf{g} \quad (7.2)$$

Since the vehicle is perfectly level ($\theta = \phi = 0$), then referring to Figure 7.9 the north and east components of acceleration are:

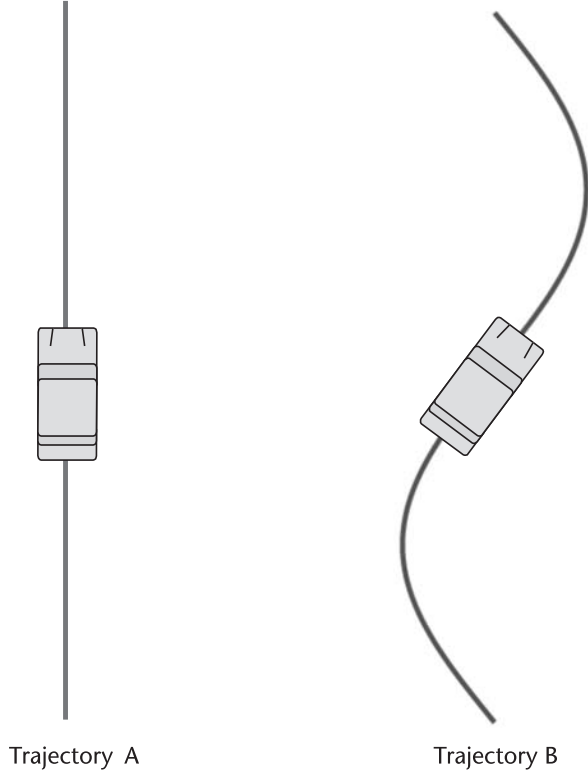


Figure 7.8 Hypothetical vehicle trajectories.

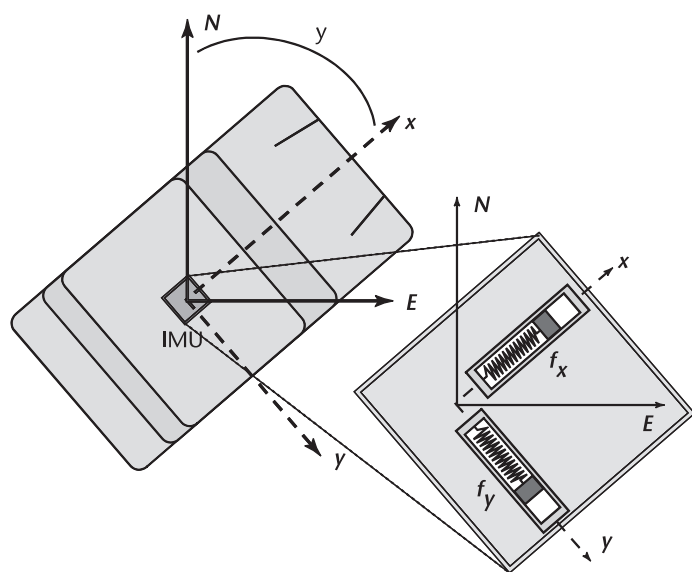


Figure 7.9 Sensor axes arrangement relative to the navigation frame.

$$a_N = \dot{v}_N = f_x \cos \psi - f_y \sin \psi \quad (7.3)$$

$$a_E = \dot{v}_E = f_x \sin \psi + f_y \cos \psi \quad (7.4)$$

For the moment, if we assume the only imperfect sensor is the gyro measuring rotation about the body z -axis of the vehicle, then the heading angle estimated by the inertial sensors, $\hat{\psi}$, will be related to the true heading, ψ , by:

$$\hat{\psi} = \psi + \delta\psi = \psi + \int_{t_0}^t b_{g_3} d\tau \quad (7.5)$$

where b_{g_z} is the z -axis rate gyro bias. Assuming that $\delta\psi$ is small allows dropping higher order terms when we substitute the above expression for $\hat{\psi}$ into (7.3) and (7.4). This gives:

$$\hat{a}_N = a_N - (f_y \cos \psi + f_x \sin \psi) \delta\psi = a_N + \delta a_N \quad (7.6)$$

$$\hat{a}_E = a_E + (f_y \sin \psi + f_x \cos \psi) \delta\psi = a_E + \delta a_E \quad (7.7)$$

The variables \hat{a}_N and \hat{a}_E represent the acceleration quantities that will be used by the inertial navigation algorithm to compute a velocity and position estimate. If the vehicle is not experiencing acceleration, f_x and f_y will be zero. Thus, $\hat{a}_N = a_N$ and $\hat{a}_E = a_E$, meaning that any heading errors resulting from gyro drift will not be reflected in the velocity and position estimates generated by using the inertial sensor outputs. This means that there will be no error signal (i.e., no difference between the GNSS and INS estimated states), which means that $\delta\mathbf{y}_k$ in (6.32) is zero and, thus, no corrections will be computed (i.e., $\delta\mathbf{x}_k = 0$). In this case, the heading error can grow without check.

Of course, any realistic maneuver will involve even slight accelerations. Thus, heading errors will *always* affect the velocity and position estimates. However, if the accelerations are small the EKF error signal $\delta\mathbf{y}_k$ will be small and very few corrections will be calculated. So, if the rate gyro biases are large, this can lead to heading errors that grow at a rate that cannot be arrested by the GNSS measurements. This, in turn, can lead to EKF divergence (i.e., errors growing without bounds and/or state estimates that do not converge to the correct value).

While the two examples above rely on low-cost inertial sensors, the following case study considers high-accuracy applications that often use higher quality inertial sensors.

7.4 Case Study 3: INS To Aid High-Accuracy GNSS

For many applications, the meter-level positioning accuracy obtained when using GNSS pseudorange measurements is insufficient. These applications rely instead on carrier-phase measurements, which, as will be discussed in Chapter 14, provide highly precise ranging capability. If the carrier phase ambiguities can subsequently

be resolved to their integer values, positioning accuracy at the centimeter level is possible. Unfortunately, resolving the ambiguities to their integer values is not a trivial task. This case study looks at how inertial navigation can help to improve the speed and reliability of the ambiguity-resolution process. However, as will be shown later, this requires an inertial navigation solution with exceptional accuracy. The low-cost inertial sensors and associated mechanizations that have thus far been the focus of this chapter will not satisfy the performance required for this application.

7.4.1 GNSS Ambiguity-Resolution Overview

A full understanding of ambiguity resolution is beyond the scope of this case study, but a cursory overview will help readers to understand the fundamental concepts. In this spirit, we consider a GNSS-only Kalman filter that uses double-difference (see Chapter 14 for details on double differencing) pseudorange and carrier-phase measurements to estimate the position states and the double-difference carrier-phase ambiguities. We write the state vector as:

$$\mathbf{x}^T = [\mathbf{p}^T \quad \mathbf{N}^T] \quad (7.8)$$

where \mathbf{p} is the position vector and \mathbf{N} is the vector of double difference ambiguities. The corresponding covariance matrix of the above states is denoted in hyper-matrix form as:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{p,p} & \mathbf{P}_{p,N} \\ \mathbf{P}_{N,p} & \mathbf{P}_{N,N} \end{bmatrix} \quad (7.9)$$

where $\mathbf{P}_{a,b}$ is the covariance matrix between vectors a and b . The covariance matrix of the ambiguities, $\mathbf{P}_{N,N}$, plays an important role in ambiguity resolution. Specifically, it has been shown in [21] that $\mathbf{P}_{N,N}$ can be used to compute the ambiguity DOP (ADOP), which is related to the volume of the hyperellipsoid defined by the ambiguity covariance matrix (centered at \mathbf{N}). The volume of the hyperellipsoid is important because it determines the size of the *search space* within which the true ambiguity vector is assumed to lie. All possible ambiguity combinations within the search space are then tested to select the best set, which is then considered as correct. Generally speaking, the smaller the search space, the faster and more reliable the ambiguity resolution process will be.

With this in mind, the GNSS-only approach to ambiguity resolution *conceptually* uses the pseudorange measurements to estimate the position. The carrier-phase measurements then use the estimated position to derive an estimate of the ambiguities with a commensurate level of accuracy. As the filter receives more GNSS measurement updates, the position estimates get better determined (until steady state), which reduces the uncertainty in the estimated ambiguities, thus reducing the search space and improving ambiguity resolution performance. Stated differently, the reason that the instantaneous ambiguity resolution is difficult is because of the inaccuracy of the pseudorange measurements relative to the carrier-phase wavelength.

7.4.2 Benefits of INSs to Ambiguity Resolution

The use of an INS can benefit the ambiguity-resolution process in two ways. First, in a tight integration configuration, the INS reduces the amount of process noise in the system model compared to the GNSS-only case. In turn, this allows for a more rapid convergence of the filter and lower steady-state covariance. This benefit is effectively realized continuously because of the nature of how the system is set up and is therefore not discussed further. The second benefit of the INS occurs following an interruption (outage) of GNSS measurements—for example, after driving through a tunnel. This benefit can be realized using a loose or a tight integration approach [22, 23] and is discussed in more detail here.

After a GNSS data outage, the carrier-phase ambiguities estimated in the Kalman filter must be reset and the convergence process described in Section 7.4.1 be restarted. Conceptually then, in a GNSS/INS system the inertial position provides additional information that is unavailable with GNSS alone, and one would therefore expect an improvement in ambiguity-resolution performance. To this end, investigations into the benefit of INS for high-accuracy positioning include, for example, [22, 27]. From [26, 27], assuming that only carrier-phase measurements are used to update the filter (for mathematical simplicity only), the ambiguity covariance matrix, $\mathbf{P}_{N,N}$, will take the following form after the first GNSS update following a GNSS outage

$$\mathbf{P}_{N,N} = \mathbf{P}_{N,N} - \lambda^2 \mathbf{P}_{N,N} (\mathbf{D}^T \mathbf{H} \mathbf{P}_{p,p} \mathbf{H}^T \mathbf{D}_{N,N} + \lambda^2 \mathbf{P}_{N,N} + \mathbf{R}_\phi)^{-1} \mathbf{P}_{N,N} \quad (7.10)$$

where λ is the wavelength of the carrier-phase measurement, \mathbf{H} is the design matrix for the undifferenced GNSS measurements, \mathbf{D} is the double difference coefficient matrix, and \mathbf{R}_ϕ is the covariance matrix of the carrier phase measurements. We draw attention to the first term in brackets, which is proportional to the position covariance matrix following the outage. The larger the covariance (in a matrix sense), the larger will be the term being inverted, and consequently, the larger the term being subtracted from the initial ambiguity covariance matrix (which is usually initialized to be large in order to reflect the unknown value of the ambiguities). Herein lies the benefit of the inertial data. Specifically, if the INS can bridge the GNSS data outage with sufficient accuracy then the ambiguity covariance matrix can be reduced relative to the GNSS-only case, resulting in improved ambiguity resolution. Although there is no hard and fast rule for defining “sufficient accuracy” in this context, [22, 23] show that once the INS accuracy degrades near that of the GNSS code solution accuracy, little benefit is gained.

Finally, it is noted that the above development said nothing of the quality of the IMU required. However, it is realized that higher quality sensors will inherently provide better INS accuracy over a given time interval (GNSS outage duration). Alternatively, for a higher quality IMU the duration over which GNSS data is unavailable can be extended relative to the case when a lower cost unit is used, with the same benefit.

7.5 Software Examples

One cannot appreciate the subtleties associated with GNSS/INS integration and inertial-aided carrier phase differential GNSS integer ambiguity resolution from the foregoing simple discussion. Thus, we urge the reader interested in gaining a deeper understanding of this topic to refer to the various papers listed at the end of this chapter. We also encourage the readers to start exercising the MATLAB-based simulations that accompany this book.

The MATLAB examples that are provided are:

1. An extended Kalman filter implementation of a loose GNSS/INS integration discussed in Section 7.2. Data from an actual IMU is provided. The reader can play back this data set and explore various aspects of GNSS/INS integration.
2. A 2D GNSS/INS integrated extended Kalman filter that demonstrates the observability analysis discussed in Section 7.3.
3. ADOP analysis examples to show the performance required by inertial navigation algorithms if they are to be used in carrier phase integer ambiguity aiding.

References

- [1] Groves, P. D., *GNSS, Inertial and Multisensor Integrated Navigation Systems*, Norwood, MA: Artech House, 2008.
- [2] Gebre-Egziabher, D., “RPV/UAV Surveillance for Transportation Management and Security,” University of Minnesota Center for Transportation Studies, *ITS Institute Report CTS 08-27*, December 2008.
- [3] Cohen, C. E., *Attitude Determination Using GPS*, Ph.D. Thesis, Department of Aeronautics and Astronautics, Stanford University, Stanford, CA, 1992.
- [4] Gebre-Egziabher, D., R. C. Hayward and J. D. Powell, “Design of Multi-Sensor Attitude Determination Systems,” *IEEE Journal of Aerospace Electronic Systems*, Vol. 40, No. 2, 2004, pp. 627–643.
- [5] Kornfeld, R., R. J. Hansman, and J. Deyst, “Single Antenna GPS-Based Aircraft Attitude Determination,” *Navigation: Journal of the Institute of Navigation*, Vol. 45, No. 1, 51–60, Spring 1998.
- [6] Gebre-Egziabher, D., and G. H. Elkaim, “MAV Attitude Determination by Vector Matching,” *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 44, No 3, July 2008, pp. 22–33.
- [7] Tseng, H. E., et al., “Technical Challenges in the Development of Vehicle Stability Control System,” *Proceeding from the 1999 IEEE International Control Conference on Control Applications*,
- [8] Hac, A., and M. Simpson, “Estimation of Vehicle Side Slip Angle and Yaw Rate,” SAE Paper 200001-0696. [36]
- [9] Nishio, A., et al., “Development of Vehicle Stability Control Based on Vehicle Sideslip Angle Estimations,” SAE paper No. 2001-01-0137.
- [10] Van Zanten, A. T., “Evolution of Electronic Control Systems for Improving the Vehicle Dynamic Behavior,” *Proceedings of the 2002 AVEC*, pp. 7–15.
- [11] Farrelly, J., and P. Wellstead, “Estimation of Vehicle Lateral Velocity,” *Proceedings from the 1996 IEEE Conference on Control Application*.

- [12] Bevy, D. M., J. C. Gerdes, and C. Wilson, C., "The Use of GPS Based Velocity Measurements for Measurement of Sideslip and Wheel Slip," *Vehicle System Dynamics*, 38(2), pp. 127–147.
- [13] O'Brien, R., and K. Kiriakidis, "A Comparison of H_∞ with Kalman Filtering in Vehicle State and Parameter Identification," *Proceedings of the American Control Conference* 2006.
- [14] Daily, R., W. Travis, and D. M. Bevy, "Cascaded Estimators to Improve Lateral Vehicle State and Tire Parameter Estimates," accepted for publication in the *International Journal of Vehicle Autonomous Systems*.
- [15] Shao, Y., and D. Gebre-Egziabher, "Stochastic and Geometric Observability of Aided Inertial Navigators," *Proceedings of the ION-GNSS 2006 Conference*, Ft. Worth, TX, Sept. 2006, pp. 2723–2732.
- [16] Bageshwar, V. L., et al., "Stochastic Observability Test for Discrete Time Kalman Filters," to appear in *AIAA Journal of Guidance, Control and Dynamics* in 2009.
- [17] Bar-Itzhack, I. Y., and N. Berman, "Control Theoretic Approach to Inertial Navigation Systems," *AIAA Journal of Guidance, Control, and Dynamics*, Vol. 11, No. 3, 1988, pp. 237–245.
- [18] Bryson, A. E., and Y. C. Ho, *Applied Optimal Control*.
- [19] Goshen-Meskin, D., and I. Y. Bar-Itzhack, "Observability Analysis of Piece-Wise Constant Systems—Part I: Theory," *IEEE Transactions on Aerospace Electronic Systems*, Vol. 28, No. 4, 1992, pp. 1065–1067.
- [20] Goshen-Meskin, D., and I. Y. Bar-Itzhack, "Observability Analysis of Piece-Wise Constant Systems—Part II: Application to Inertial Navigation In-Flight Alignment," *IEEE Transactions on Aerospace Electronic Systems*, Vol. 28, No. 4, 1992, pp. 1068–1075.
- [21] Rhee, I., M. Abdel-Hafez, and J. Speyer, "Observability of Integrated GPS/INS During Manuevers," *IEEE Transactions on Aerospace Electronic Systems*, Vol. 40, No. 2, 2004, pp. 526–535.
- [21] Teunissen, P. J. G., and D. Odijk, "Ambiguity Dilution Of Precision: Definition, Properties, and Application," *Proceedings of ION GPS 1997*, The Institute Of Navigation, pp. 891–899, 1997.
- [22] Petovello, M. G., *Real-Time Integration of a Tactical-Grade IMU and GPS for High-Accuracy Positioning and Navigation*, Ph.D. dissertation, Department of Geomatics Engineering, University of Calgary, Calgary, AB, p. 242.
- [23] Petovello, M. G., M. E. Cannon, and G. Lachapelle, "Benefits of Using a Tactical-Grade IMU for High-Accuracy Positioning," *Navigation: Journal of the Institute of Navigation*, The Institute Of Navigation, Vol. 51, No. 1, pp. 1–12, 2004.
- [24] Škaloud, J., *Optimizing Georeferencing of Airborne Survey Systems by INS/DGPS*, Ph.D. dissertation, Department of Geomatics Engineering, University of Calgary, Calgary, AB, p. 160.
- [25] Škaloud, J., "Reducing The GPS Ambiguity Search Space By Including Inertial Data," *Proceedings of ION GPS 1998*, The Institute Of Navigation, pp. 2073–2080, 1998.
- [26] Scherzinger, B. M., "Robust Positioning with Single Frequency Inertially Aided RTK," *Proceedings of ION NTM*, The Institute Of Navigation, pp. 911–17, 2002.
- [27] Scherzinger, B. M., "Precise Robust Positioning with Inertially Aided RTK," *Navigation: Journal of the Institute of Navigation*, The Institute Of Navigation, Vol. 53, No. 2, pp. 73–83, 2006.

Integrated LADAR, INS, and GNSS Navigation

Jacob L. Campbell and Maarten Uijt de Haag

8.1 Introduction

In this chapter we examine another integrated navigator that fuses information derived from terrain databases, laser detection and ranging or laser radar (LADAR) mapping systems, inertial navigators, and GNSSs to realize a multisensor system for aerial navigation applications. As shown in Table 8.1, systems that use a terrain database and radar to estimate position updates for a navigation system have been in existence since the 1950s. However, with the advent of the GNSS (more specifically the GPS), terrain-referenced navigation systems have become less prominent given the typically superior performance and worldwide availability of GNSS. In general, the accuracy of a terrain-referenced navigation system is limited to the accuracy and resolution of the “referenced” terrain databases as well as the accuracy and resolution of the sensor onboard the aircraft that measures the shape of the terrain. With the advent of new terrain mapping technologies, specifically light detection and ranging (LIDAR) mapping systems, terrain databases are currently being created and stored at resolutions better than a meter (horizontal) with decimeter-level accuracies. This milestone, coupled with the developments of commercial off-the-shelf (COTS) airborne laser scanners and efficient processing techniques, has allowed terrain-referenced navigation to exceed some aspects of GPS performance for various airborne applications.

Terrain-referenced navigation systems possess a desirable characteristic found in many dead reckoning systems: the ability to operate independently of terrestrial or space-based radio-navigation aids after initialization. However, because these systems use a georeferenced terrain database, they do not suffer from the continuous error growth found in traditional dead reckoning systems such as INSs. Terrain-referenced navigation systems typically consist of the following three components: a ranging sensor such as radar or LADAR that measures the shape of the terrain traversed by the aircraft; a sensor that estimates the aircraft’s trajectory while accumulating terrain shape measurements such as a baro-aided INSs; and a georeferenced terrain database.

This chapter is organized into three main parts. First, a general background of terrain-aided inertial navigation is given. This will include details on one such laser-based terrain-referenced navigation system—the terrain-aided inertial navigator (TERRAIN). The second part discusses the mathematics of terrain-aided

navigation algorithms such as the ones found in TERRAIN. The final part of the chapter provides two case studies highlighting the performance of TERRAIN. Readers interested in more detail on the subject should refer to Appendixes 8A and 8B, which, respectively, discuss the historical background of terrain-aided navigation and provide a survey of actual fielded terrain-aided navigators. Table 8.1 lists the systems covered in this survey. These appendixes can be found on the DVD that accompanies this book.

8.2 LADAR-Based TERRAIN Integration Methodology

In general, the accuracy of a terrain-referenced navigation system is determined by the accuracies of both the terrain database and the sensor that is used to make

Table 8.1 Terrain-Aided Navigators Discussed in Appendix 8B (*System Accuracies Based on Information Contained in References Given in Associated Sections in Appendix 8B)

<i>TRN System</i>	<i>Terrain Sensor</i>	<i>Period of Use</i>	<i>Application</i>	<i>Accuracy*</i>	<i>Terrain Data</i>	<i>Comments</i>
ATRAN section (more details in [10])	X-Band horizontal scanning pulse radar	Early 1950s through 1960	MACE cruise missile	305m with 152-m repeatability	Created on 35-mm film using scale terrain models	Analog system, weight 1200 lbs
TERCOM Section 8B.1	C-Band Downward- looking Rad Alt	1970s to present	Tomahawk cruise missile	30.5-m horizontal CEP	122-m postsampling (1974 figure)	Batch processing TRN system
SITAN Section 8B.2	C-Band downward- looking Rad Alt	1970s to 1980s	Prototype trials on Aircraft	Between 19- and 75-m horizontal CEP	Unpublished	Sequential processing TRN system
SPARTAN Section 8B.3	C-Band Downward- Looking Rad Alt	1980s	Prototype trials on aircraft	Unpublished	Unpublished	Processing based on Bayesian statistics
TERPROM Section 8B.4	C-Band downward- looking Rad Alt	1980s to present	Aircraft and Cruise Missile	30-m horizontal CEP 5-m vert. LEP	Unpublished	Widely used in military aircraft
APALS Section 8B.5	X-Band forward, side- to-Side scanning Wx radar	1990s to present	Prototype aircraft landing system	2–3 m horizontal 1–2m vertical	Spotlight SAR image map	Currently Delayed in Certification process
PTAN Section 8B.6	Interferometric C-Band downward- looking Rad Alt	1990s to present	Airborne applications	3-m horizontal (<5,000m AGL) 30-m horizontal (from 5,000 to 30,000m AGL)	DTED Level 4 (3-m postsampling)	Along track Doppler- window resolution dependent on vehicle velocity

observations of the terrain. Therefore, an improved terrain-referenced navigator design should be based on the combination of a higher resolution, higher accuracy terrain database and a higher update rate, higher accuracy sensor. The technology that has enabled much of this research is the airborne LIDAR mapping system which includes a LADAR sensor. The recent proliferation of LIDAR mapping systems in the mapping community has made available high-resolution (1-m horizontal measurement spacing or better) and accurate (elevation measurement accuracies at 30-cm RMS [1]) terrain models. The TERRAIN system described in this chapter estimates the position of an aerial vehicle by integrating measurements from an IMU with those of a LADAR, and a high-resolution/high-accuracy (1-m resolution and 30-cm RMS elevation accuracy) terrain database or digital elevation model (DEM) created using a LIDAR. While desirable, a LADAR-based system is currently limited by existing laser technology to clear-weather and/or low-altitude applications. However, as with many technologies, research in the field of lasers is enabling longer wavelengths, variable power, and sophisticated laser receiver-processing techniques at lower cost—all of which are necessary in expanding the operational envelope of LADAR systems. While the technology may not be ready for use in the civil aviation community, advances in these areas will increase its desirability for navigation. The capability to provide positioning estimates without navigation aids is of interest not only to the civil aviation community, but also for UAVs, which can operate in urban environments where the performance of radio-navigational aids is degraded, unavailable, or purposely denied.

LADAR-based systems have the potential to be used in aircraft applications beyond positioning, including attitude determination, runway incursion detection, and obstacle detection [2, 3]. Depending on the required performance for these applications, LADAR would be integrated with other on-board sensors or map information. For en route position and attitude determination, integration of LADAR with an INS and a DEM is envisioned. For general positioning capability, LADAR plus a DEM may be sufficient. For both runway incursion and obstacle detection, LADAR with INS may suffice. This chapter is concerned primarily with the performance of LADAR integrated with INS and a DEM for aircraft positioning. When initialized with the GPS WAAS, TERRAIN may be able to provide horizontal and vertical guidance for an aircraft-precision approach.

Various challenges must be overcome before LADAR terrain-referenced positioning technology can be used for all-weather LADAR-based aircraft navigation. One challenge is operation of the laser range scanner under various weather conditions, such as fog, rain, cloud cover, and snow. Three possible solutions are to increase the power transmitted by the laser, to increase the sensitivity of the detector, or to limit the range at which the laser would be operated. At the same time, the safety aspects of laser operation must be considered, as discussed in [4]. Other solutions to the weather penetration problem may include the use of multispectral laser systems and processing of the laser data with the aid of an IMU to allow for improved data correlation. A second challenge is the current cost of many of today's systems that use LADAR, such as an airborne LIDAR mapping system, which often costs more than \$1 million. However, LIDAR systems have very high accuracy requirements that are not necessary for a LADAR-based positioning system. It is of note that a LADAR-based navigator capable of operating at a maximum height

of 400m in clear weather has been implemented using a ship-borne INS and COTS LADAR and computers for less than \$150,000, with no extensive calibration process required [5].

To take full advantage of a LADAR system it is necessary to have DEMs that are accurate and reliable and have a spatial resolution comparable to the point cloud generated by the LADAR system. Such DEMs, certified for use in aircraft terrain navigation systems, are not available at this time. RTCA DO-276 provides guidance for the development of terrain and obstacle databases for several aeronautical purposes, including terrain awareness and warning systems (TAWs), synthetic vision systems, and instrument procedure design [6]. It is noted several times in [6] that specifications given “are not necessarily sufficient for primary means of navigation.” Thus, specifications for the creation of high-accuracy and high-resolution DEMs (submeter accuracy with less than 5-m postspacing) for use in aircraft navigation must be developed before a certifiable LADAR positioning system can be realized. With these specifications in place, it will then be necessary to conduct mapping missions to create an extensive collection of high-accuracy/resolution DEMs for aircraft navigation. In addition, given the overall shape of terrain does not vary greatly over time, the correctness of terrain databases used in the approach and landing segments of flight can be verified through qualification means similar to the ILS. Also of concern is the method used to create the DEM from an airborne LIDAR mapping system. Depending on the application, LIDAR-generated DEMs may be processed to create a bare-Earth model (trees and buildings removed), to include all the detected trees and buildings, or to contain some features with the terrain.

The functional components of a generic LADAR terrain-referenced navigation system are shown in Figure 8.1. The LADAR-based TERRAIN system described in this chapter is based on the various terrain-referenced navigation concepts discussed above. The TERRAIN system has been flown in real time and operates in one of two modes during the approach phase of flight: WAAS mode or TERRAIN mode. Figure 8.2 provides a block diagram of the approach system. During the initial approach phase, the TERRAIN approach system is in WAAS-aided mode. In this mode, WAAS is used to update the position estimates, and GPS-integrated carrier-phase measurements are used to calibrate the integrated velocities from the INS (mode controller = 0 in Figure 8.2). WAAS is used until the decision height (DH) specified for the WAAS approach procedure is reached, at which point the mode is switched to TERRAIN mode (mode controller = 1 in Figure 8.2). In TERRAIN mode, the INS velocity error estimates in the Kalman filter are fixed and the corrected INS integrated velocities are added to the most recent position estimate to form the corrected position prediction. The corrected position predictions are used in the terrain navigator to transform the points measured by the LADAR to a georeferenced frame. After the generation of the point cloud, the terrain position estimator is used to estimate the error in the current corrected position. This estimated error is fed back into the terrain position estimator for inclusion in the next position estimate as shown in Figure 8.2. The next two sections provide an overview of the terrain-referenced position estimation and the WAAS-aided inertial Kalman filter.

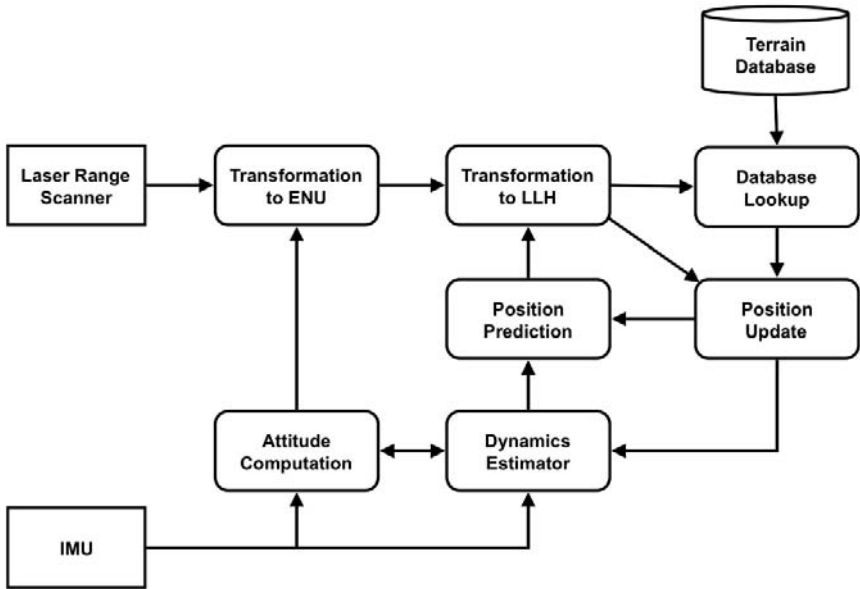


Figure 8.1 Block diagram of inertial-based navigation system with LADAR aiding.

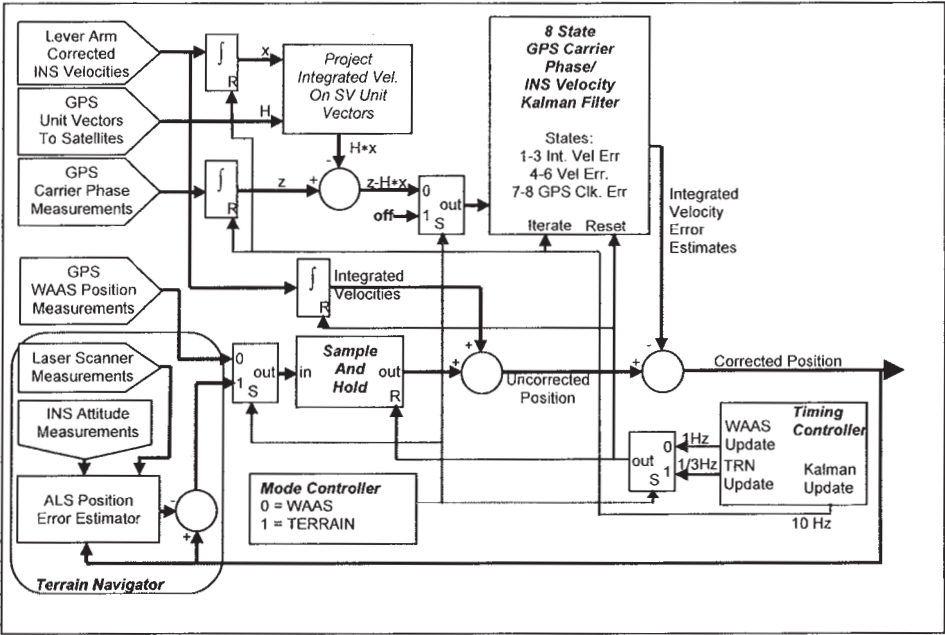


Figure 8.2 TERRAIN precision approach system position estimator.

8.3 LADAR-Based Terrain-Referenced Position Estimation

The proposed LADAR terrain-referenced positioning system shown in the left, bottom block in Figure 8.2 is based on the calculated agreement between the terrain signature derived from LADAR sensor data and the terrain signature retrieved from the DEM stored onboard the aircraft. As noted earlier, the basic functional components of the LADAR terrain-referenced navigation system are shown in Figure 8.2. The implemented aircraft position estimator minimizes the sum-of-squared-error (SSE) between both terrain signatures. Two methods have been used to solve for the minimum SSE: an exhaustive search method and a gradient-based search method. A more detailed description of the dynamics estimator and the attitude computation implementations depicted in Figure 8.2 are provided in [7].

8.3.1 Position Estimate and SSE Surface

To calculate the position update function, measurements from the LADAR are compared with data in the DEM. To perform this comparison, the LADAR measurements must be converted to a set of elevation values expressed in the DEM coordinate frame. Thus, a set, S , of number, N , LADAR range measurements, $S_{\text{ALS}} = \{t_i, \alpha_i, \rho_i \mid i = 1, \dots, N\}$, where t_i = time, α_i = laser scan angle, and ρ_i = laser range, are converted from the aircraft body frame to a point cloud in an east-north-up (ENU), local-level coordinate frame identified by the set, $S_{\text{ENU}} = \{t_i, \mathbf{x}_{\text{ENU}}(t_i) \mid i = 1, \dots, N\}$. The origin of the local level coordinate frame coincides with the laser scanner optics reference point (LSO) at time t_0 . The transformation from the body to the ENU coordinate frame requires attitude estimates of the LADAR (pitch, θ , roll, ϕ , and heading, ψ), and integrated LADAR velocities starting at t_0 . Estimated incremental ENU frame rotations can be included for completeness; however, for this application, the effect the ENU frame rotations have on the generation of the ENU reference point cloud is negligible over the period of one second at airplane velocities. The attitude estimates are generated in the attitude computer based on measurements from the IMU (for example, using (6.8) described in Section 6.2 or similar methods described in [7]). The integrated velocity estimates are computed as follows:

$$\hat{\mathbf{x}}_{\text{LSO, ENU}}(t_i) = \hat{\mathbf{x}}_{\text{LSO, ENU}}(t_0) + \int_{t_0}^{t_i} \hat{\mathbf{v}}_{\text{LSO, ENU}}(t) dt \quad (8.1)$$

where the velocity estimates, $\hat{\mathbf{v}}_{\text{LSO, ENU}}(t)$, are calculated in the dynamics estimator and based on IMU velocity measurements.

The local-level point cloud coordinates can then be converted to World Geodetic Survey 1984 (WGS-84) latitude, longitude, height (LLH) coordinates defined by the set,

$$S_{\text{LLH}} = \{t_i, \hat{\mathbf{x}}_{\text{LLH}}(t_i) = [\hat{L}(t_i) \quad \hat{\lambda}(t_i) \quad \hat{h}_{\text{ALS}}(t_i)]^T \mid i = 1, \dots, N\}$$

The ENU-to-LLH coordinate transformation at time-epoch t_i is a function of the predicted position of the laser scanner optics at the time-epoch t_i . The corresponding terrain elevation, $h_{\text{DEM}}(\hat{L}(t_i), \hat{\lambda}(t_i))$ at the estimated latitude coordinate, $\hat{L}(t_i)$, and longitude coordinate, $\hat{\lambda}(t_i)$, of S_{LLH} , is retrieved from the DEM server using bilinear interpolation of the four closest DEM points given that the DEM is stored as a uniformly spaced grid of elevations.

The agreement metric is defined as the difference between the measured height and the height synthesized from the DEM. This difference or disparity can be computed for each laser range measurement. Thus, the set, P , of N disparities can be obtained:

$$P = \{p(t_i) = \hat{h}_{\text{ALS}}(t_i) - h_{\text{DEM}}(\hat{L}(t_i), \hat{\lambda}(t_i)) \mid i = 1, \dots, N\} \quad (8.2)$$

The SSE can now be evaluated using the set of N disparities, or:

$$\text{SSE}_N = \sum_{i=1}^N \{p(t_i)\}^2 \quad (8.3)$$

Both a grid-based and a gradient-based method can be used to find the position offset for which the SSE_N value is minimum, resulting in an aircraft position estimate. In case of the grid-based method, a search area of known position offsets will be defined around the position predictions determined by (8.1). The size of this search area will be determined by the covariance of the position predictions. The predicted aircraft position at each time-epoch, t_i , is now replaced by a set of candidate aircraft positions:

$$\hat{\mathbf{x}}'_{\text{LSO,ENU}}(t_i, j, k, m) = \hat{\mathbf{x}}_{\text{LSO,ENU}}(t_i) + \Delta \hat{\mathbf{x}}_{\text{LSO,ENU}}(t_i, j, k, m) \quad (8.4)$$

where

$$\Delta \hat{\mathbf{x}}_{\text{LSO,ENU}}(t_i, j, k, m) = \begin{bmatrix} j \cdot \Delta x \\ k \cdot \Delta y \\ m \cdot \Delta h \end{bmatrix} \quad (8.5)$$

for $j = -M, \dots, M$, $k = -M, \dots, M$, and $m = -M, \dots, M$. Δx and Δy determine the resolution of the horizontal dimensions of the search area. For clarity, the search will be divided into two searches: a horizontal search and a vertical search. For a horizontal search, (8.5) simplifies to $\Delta \hat{\mathbf{x}}_{\text{LSO,ENU}}(t_i, j, k) = [j \cdot \Delta x \quad k \cdot \Delta y \quad 0]^T$ and the disparity in (8.2) can now be evaluated at each point $\hat{\mathbf{x}}'_{\text{LSO,ENU}}(t_i, j, k)$, or

$$P(j, k) = \{p(t_i, j, k) = h_{\text{ALS}}(t_i) - h_{\text{DEM}}(t_i, j, k) \mid i = 1, \dots, N\} \quad (8.6)$$

Similarly, the SSE metric can be evaluated for all values of j and k , or

$$SSE_N(j, k) = \sum_{i=1}^N \{p(t_i, j, k)\}^2 \quad (8.7)$$

The set of $M \cdot M$ points is referred to as the SSE surface. The horizontal position estimate, or position of best agreement, can now be obtained by finding the values of j and k for which the SSE_N value is a minimum. Mathematically this can be expressed as follows:

$$(\hat{j}, \hat{k}) = \arg \min_{j, k} \{SSE_N(j, k)\} \quad (8.8)$$

The corresponding position estimate can be derived from:

$$\hat{\mathbf{x}}_{\text{LSO, ENU}}(t_i) = \begin{bmatrix} \hat{x} + \hat{j} \cdot \Delta x \\ \hat{y} + \hat{k} \cdot \Delta y \\ \hat{h} \end{bmatrix} \quad (8.9)$$

where \hat{h} is the height estimate that can be computed by extending the search space to the third dimension. In the system described in this chapter, both search techniques are used to find the optimal values for j and k : an exhaustive grid search and a gradient-based search.

8.3.2 Exhaustive Grid Search

The exhaustive grid search algorithm entails the evaluation of $SSE_N(j, k)$ for each value of j and k within the search grid. To demonstrate this process, two sets of exhaustive searches were performed on 60s of LADAR data collected during a flight test on NASA Dryden's DC-8. The first exhaustive search was over an area of $\pm 30\text{m}$ in the east/west and north/south directions with a 1-m step interval ($\Delta x = 1\text{m}$, $\Delta y = 1\text{m}$) and a $\pm 4\text{-m}$ search around the aircraft's a priori height (the up direction) with a 30-cm step interval. An example of one of the horizontal search spaces is shown in Figure 8.3.

The second exhaustive search was performed in the east/west and north/south direction with a $\pm 9\text{-m}$ search area and a 30-cm step interval and a search area of $\pm 4\text{ m}$ in the up direction with a 30-cm step interval. Figure 8.4 shows the results of this horizontal search.

Each of the exhaustive search methods, depicted in Figures 8.3 and 8.4, uses one second of LADAR data. At a pulse repetition frequency (PRF) of 33,333 Hz this corresponds to $N = 33,333$ measurements for the position estimation process. For this value of N , the implementation of the exhaustive search requires over 3 billion DEM lookups and bilinear interpolations. The large number of DEM lookups performed by this method takes over 23 min on a Pentium 4, 1.8 GHz PC with 1.3 GB of RAM. Although the exhaustive search provides a good insight into the characteristics of the SSE surface, alternative methods for finding the SSE surface minimum must be sought for the implementation of real-time LADAR terrain-referenced navigation.

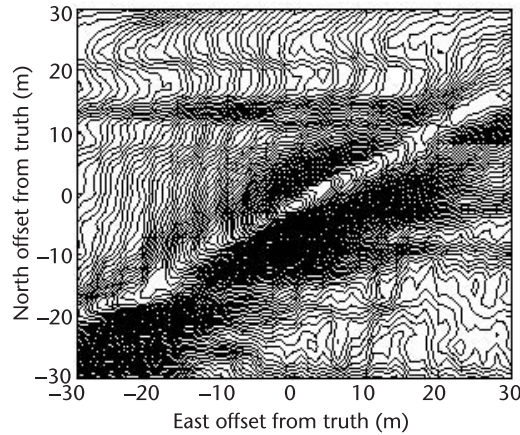


Figure 8.3 SSE surface for GPS time 314,246s of week 1229 (30×30 -m search area, 1-m spacing).

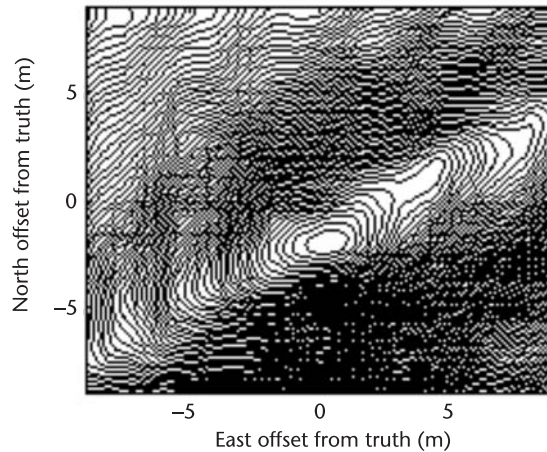


Figure 8.4 SSE surface for GPS time 314,246s of week 1229 (9×9 -m search area, 0.3-m spacing).

Figure 8.3 shows that more than one local minimum exists on the SSE surface. However, as the “area of interest” is reduced from 30-by-30m (Figure 8.3) to 9-by-9m (Figure 8.4) the number of local minima is reduced to one. This characteristic was found to exist over the 60s of SSE surfaces examined. Therefore, the following supposition was made: If the predictions in (8.2) are sufficiently good (better than 9m) using a search to find the local minimum should yield the global minimum on the SSE surface, thus yielding the highest terrain agreement. This supposition allows for a significant reduction in the search for the SSE minimum with respect to the exhaustive search with over one hundred thousand SSE calculations, each with 33,333 measurements, to typically less than 100 SSE calculations with 33,333 measurements each. Section 8.3.3 presents the use of a gradient search to find the minimum on the SSE surface.

8.3.3 Gradient-Based Search

A reduction in required computations can be achieved by using gradient-based techniques to search for the SSE surface minimum given the assumption that there is only one minimum in the search area. The gradient at a point on a discrete grid surface can be approximated to the first order by differencing the data around that point. Many methods exist to approximate the gradient at specific locations on a grid-defined surface; for an overview of these gradient approximation methods the reader is referred to [8]. However, to define the gradient on the SSE surface some assumptions must be made. First, care should be taken in choosing an appropriate value for the distance from the position at which the gradient is calculated to the four points on the SSE surface used to compute this gradient. Several factors, which should be considered when choosing this parameter, are discussed below. Second, it is assumed that the LADAR measurements are spatially and temporally independent (i.e., the laser illumination points do not overlap, and the corresponding points in the DEM are independent). The size of the uniform grid postspacing in the DEM is small enough such that each laser measurement uses an independent set of DEM heights to calculate the disparity. It is also assumed that the LADAR measurements are uniformly and randomly distributed within the area defined by the four closest DEM heights since bilinear interpolation is used to compute a DEM height of a given position. The SSE surface is not defined by an explicit function; hence, empirical results were obtained and used to identify the parameters required to estimate the SSE surface gradient for the specific case of the flight test described later in this chapter.

In case of the gradient-based search, (8.4) is modified as follows:

$$\hat{\mathbf{x}}_{\text{LSO,ENU}}(t_i)|_k = \hat{\mathbf{x}}_{\text{LSO,ENU}}(t_i)|_{k-1} + m_{k-1} \Delta \mathbf{x}_{\text{LSO,ENU}}|_{k-1} \quad (8.10)$$

where $\cdot|_k$ indicates the k th iteration in the search, m_k is the step size at iteration k , and $\Delta \mathbf{x}_{\text{LSO,ENU}}|_{k-1}$ is found by:

$$\Delta \mathbf{x}_{\text{LSO,ENU}}|_k = -\frac{1}{\|\nabla \text{SSE}_N|_k\|} \begin{bmatrix} \frac{d\text{SSE}_N|_k}{dx} & \frac{d\text{SSE}_N|_k}{dy} & 0 \end{bmatrix}^T \quad (8.11)$$

where $\|\nabla \text{SSE}_N|_k\|$ is the gradient of the SSE surface evaluated at $\hat{\mathbf{x}}_{\text{LSO,ENU}}(t_i)|_k$ and $d\text{SSE}_N|_k/dx$ and $d\text{SSE}_N|_k/dy$ are the individual gradient components at $\hat{\mathbf{x}}_{\text{LSO,ENU}}(t_i)|_k$.

The algorithm can now be described as follows.

- For time epoch $k = 0$:
 - $\hat{\mathbf{x}}_{\text{LSO,ENU}}(t_i)|_0$ = initial position estimate of the laser scanner optics (LSO) in ENU at time t_i .
 - Compute $\Delta \mathbf{x}_{\text{LSO,ENU}}|_0$ using equation (8.9).
 - m_0 = initial step size (based on predicted INS position error growth between updates, e.g. 2 m).

- For time epochs $k = 1$ to number of iterations, M'' :
 - Update position estimate vector $\hat{\mathbf{x}}_{\text{LSO, ENU}}(t_i)|_k$ using (8.8).
 - Compute $\Delta \mathbf{x}_{\text{LSO, ENU}}|_k$ using equation (8.9).
 - m_k = distance to travel in the $\Delta \mathbf{x}_{\text{LSO, ENU}}|_k$ direction:

$$m_k = \begin{cases} m_{k-1} & \text{if } |\mathbf{x}_k - \mathbf{x}_{k-2}| \geq A \cdot m_{k-1} \text{ or } k < 2 \\ 0.5m_{k-1} & \text{if } |\mathbf{x}_k - \mathbf{x}_{k-2}| < A \cdot m_{k-1} \end{cases}$$

where A = travel distance reduction decision constant (set to 1.5m for the system described later in this chapter).

As described in the above algorithm, an estimated SSE surface gradient is used to determine the direction of the next step in the gradient search. The distance traveled in each step is controlled by another mechanism that reduces this distance as oscillations around a minimum are observed. The oscillating state is determined by computing the distance between the current gradient calculation position and the gradient calculation made two steps before. Figure 8.5 illustrates one example of the SSE surface gradient search.

The use of a gradient-based search for the minimum on the error surface reduces the computation time from 23 minutes for one second of data with an exhaustive search, to near real time in a Matlab implementation (i.e., a performance increase that is greater than a factor of 400). Further improvements in the implementation of the gradient search have resulted in real-time operation. These improvements include storage of the terrain database in random access memory (RAM) and the implementation of the code in C on a real-time operating system (RTOS).

The agreement between the LADAR data and the terrain database is evaluated by comparing the vertical component of points illuminated by the LADAR and the vertical component of a height retrieved from the terrain database given a LADAR horizontal position estimate. Since a direct comparison between the vertical

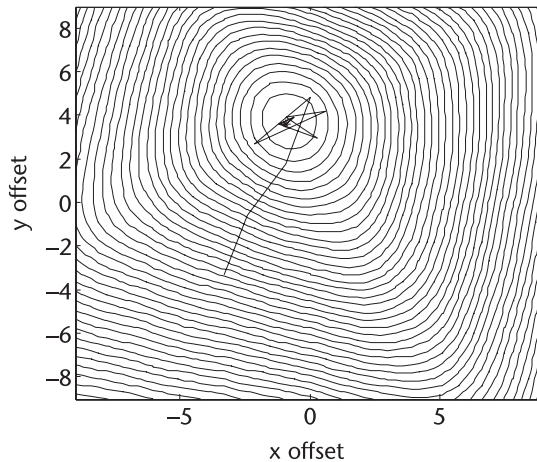


Figure 8.5 Gradient search for minimum error on the sum of squared error surface.

components is performed, small errors in the predicted horizontal position can map into a large error in the vertical if features such as buildings or trees are present. The implementation of a simple outlier removal filter reduced these effects and improved the positioning performance. The outlier filter removes points that are more than several standard deviations larger than the expected vertical error. For example, if the vertical error was computed to be approximately 1m RMS, then a 6σ value of 6m could be selected to remove measurements that were influenced by trees or buildings.

8.4 Estimation of Inertial Velocity Error

To achieve LADAR-based navigation, position estimates generated using the general methods of Section 8.3 are combined with an INS using a Kalman filter. For the case studies that will be presented in Section 8.5, the INS used was a Honeywell HG-1150. This is a navigation grade INS (see Table 6.1). The velocity error estimates for the Honeywell HG-1150 INS are obtained from a Kalman filter, while the aircraft is on initial approach and using WAAS for guidance. Upon the transition to the final approach phase, the Kalman filter is disabled and the last velocity error estimates are used to complete the TERRAIN approach. Thus, the Kalman filter is not active during the final approach phase where the TERRAIN system is used. The implementation of the Kalman filter was performed in C++, and a detailed description of the filter can be found in [9]. The observations input to the Kalman filter are the differences between the GPS integrated carrier phase for each space vehicle (SV) and the integrated inertial velocities mapped onto the line-of-sight (LOS) vectors between the estimated user position and each SV. The eight state-variable Kalman filter has the following state variables: 1 through 3, integrated velocity error in ENU; 4 through 6, velocity error in ENU; and 7 and 8, GPS clock bias and drift. As seen in Figure 8.1, the integrated velocity error estimates (variables 1 through 3) are used to remove the estimated error in the INS's integrated velocities.

8.5 Case Studies of TERRAIN System Performance

Sections 8.5.1 and 8.5.2 present case studies of the laser-based navigation concepts described in the previous sections. The two case studies that will be considered are based on two flight tests. The first set of flights (case I) were conducted on July 28, July 30, August 1, and August 4, 2003, with NASA Dryden's DC-8 Flying Laboratory as part of NASA Langley Research Center's (NASA LaRC) Aviation Safety Program. The second set of flights included eight real-time TERRAIN precision approaches to Runway 19 at Braxton County Airport (K48I), West Virginia, on January 14, 2005.

8.5.1 Case Study I—General Positioning System

The first set of flight test results demonstrated the use of the TERRAIN system as a general purpose positioning system for applications such as en route navigation.

For this test a data collection system was installed on a DC-8 and LIDAR data, kinematic GPS data, IMU attitude and velocities, laser scanner angle data, laser scanner range data, radar altimeter data, and weather radar data were collected. The LIDAR system used to collect LADAR data was an Optech, Inc. Airborne LIDAR Terrain Mapper (ALTM) on loan from the U.S. Army. The LIDAR laser unit was mounted in the cargo bay of NASA Dryden's DC-8 (Figure 8.6) Flying Laboratory, as shown in Figure 8.7.

Data was collected for approaches flown into the Reno, Nevada, airport (KRNO). The collection of the LADAR data was limited to a minimum-altitude above-ground level (AGL) to satisfy eye-safe regulations. Figure 8.8 shows a trajectory of one approach and its surrounding terrain.

Over 12 Gb of binary LIDAR/navigation data was collected during the five missions flown for the July and August DC-8 flight tests. Much of this data was collected outside the 18-by-10-km region covered by the available National Geodetic Survey (NGS) DEM. The laser scanner settings varied from 10 to 20 deg in scan width and from 15 to 29 Hz in scan rate. KRNO is in the center of the high-accuracy/resolution DEM. Flights over the DEM were made as approaches to the airport; the center region of the DEM contains no laser measurements because the altitude while descending on each approach fell below the allowed eye-safe level. In total, 277s of LADAR data was processed for this study, with the longest continuous data segment being 62s in duration. No filter techniques were applied to the laser measurements to remove laser measurement outliers (e.g., returns from clouds and birds).

The gradient search of Section 8.3 was used to calculate the position estimate for each 1-s set of data; thus 33,333 laser measurements were used for each position



Figure 8.6 NASA Dryden DC-8 flying laboratory. (Photo courtesy of NASA Dryden. Reprinted with permission.)

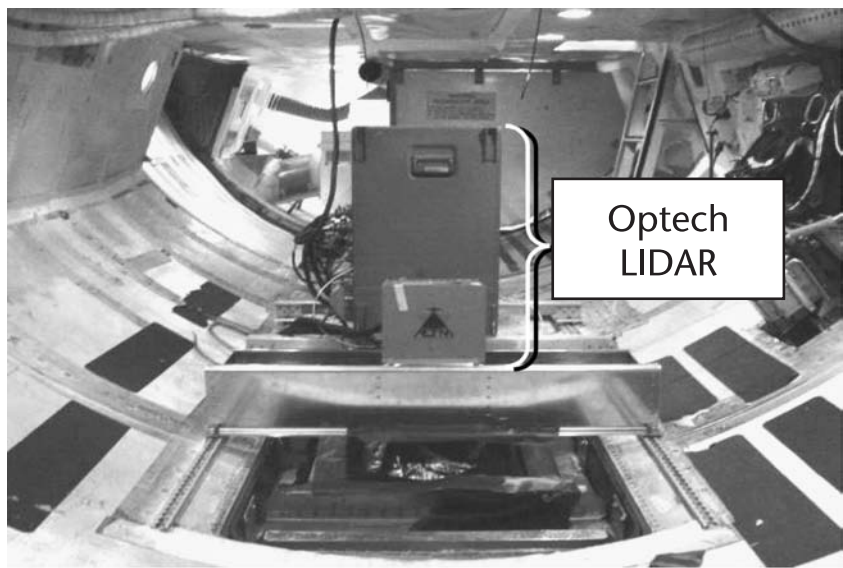


Figure 8.7 NASA Dryden DC-8 cargo bay LIDAR installation.

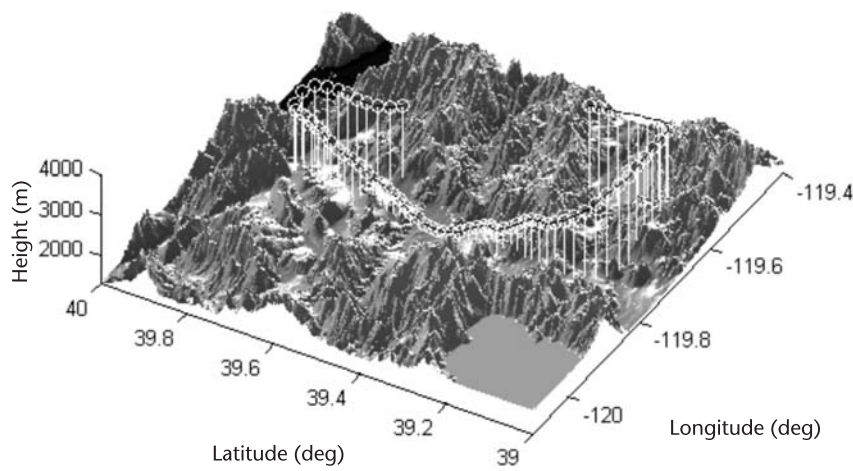


Figure 8.8 Flight path of an approach into KRNO airport.

estimate. Each gradient search was permitted to iterate a maximum of 20 times with an initial step of $m_0 = 2\text{m}$. The height error space was searched by performing several gradient searches on each 1-s piece of data for several height offsets, comparing the minimums from each search, and selecting the search that contained the highest agreement. Figure 8.9 shows a plot of the horizontal errors in the LADAR-based position estimates for all 277s of LADAR data.

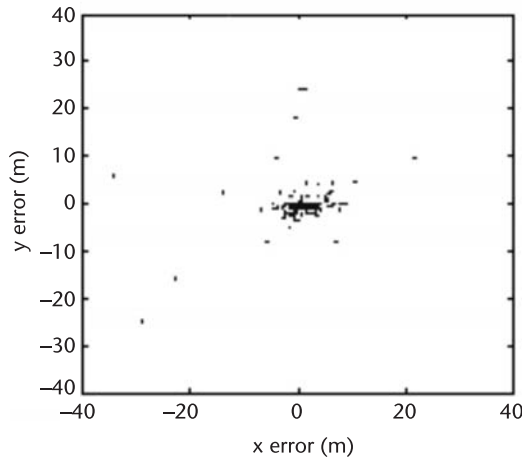


Figure 8.9 LADAR horizontal position estimate error.

The error values in the x (east/west) and y (north/south) directions were computed by subtracting the LADAR navigation position estimate from the post-processed kinematic GPS (KGPS) data (the truth reference data). Over 90 percent of the position estimates fall within 10m of the true position. Table 8.2 provides a summary of the position estimate characteristics for the 277s of data processed. Table 8.2 shows that the horizontal radial position error is 8.9m (1σ), while the vertical position error is 0.51m (1σ) for 1-s position updates. If a 10-m data-editing filter is introduced that discards solutions that disagree by more than 10m with respect to the INS position solution, the positioning accuracies improve to approximately 3.2m (1σ) and 0.5m (1σ) for the horizontal radial and vertical position errors, respectively.

8.5.2 Case Study II—Precision Approach Guidance System

Real-time TERRAIN precision approaches were performed at Braxton County Airport, Ohio, (K48I) to runway 19. The view from the DC-3 on short final can be seen in Figure 8.12. K48I was chosen because of the availability of LIDAR-derived maps and the relatively short proximity (approximately 80 nmi) to the Ohio University Airport (KUNI). Runway 19 at K48I is 4,000 ft in length and

Table 8.2 Summary of LADAR Position Estimates (1-s updates)

<i>Description</i>	<i>Mean (meters)</i>	<i>Standard Deviation (meters)</i>
x error (all 277 points)	0.45	5.12
y error (all 277 points)	-0.87	7.28
h error (all 277 points)	-1.16	0.51
x error with 10-m cutoff (251 points)	1.00	2.74
y error with 10-m cutoff (251 points)	-0.51	1.67
h error with 10-m cutoff (251 points)	-1.15	0.50

60 ft in width. Currently, there are no published approach procedures into this airport. An airport survey was conducted and the runway touchdown point for the approach was created to be 1,000 ft from the threshold and a straight-in 3-deg approach was extended from the touchdown point.

One of the most challenging aspects of the Braxton County flight test proof-of-concept demonstration was the development of a TERRAIN system that functions in real time. Given the amount of data to be processed and the characteristics of the algorithms of the real-time system, a three-computer architecture was developed as shown in Figure 8.10. The first computer is the data collection and distribution (DCD) computer. Its primary task is to time-tag, store, and distribute the data from the various sensors used in the system. The second computer is the navigation (NAV) computer, which uses a multithreaded architecture to perform the positioning and navigation functions. The third computer uses the aircraft state data, computed by the NAV computer, to render the guidance cues displayed to the pilot.

Since integrity of the data links between the computers was not an objective of the proof-of-concept system, Ethernet was used to connect the three computers. This section provides a description of the GPS WAAS, INS, and LADAR sensors connected to the DCD computer, as well as a description of the setup of the three computers used in the proof-of-concept system.

Measurements of the terrain are made with the class 1 (eye safe) Riegl LMS-Q140i airborne laser scanner. For the real-time TERRAIN precision approaches the LMS-Q140i was configured to scan with a 60-degree cross-track scan at 15 scans per second with an average PRF of 10,000 range measurements per second. Assuming an altitude of 1,000 ft AGL and a ground speed of 60 m/s, the distance between scans in the along-track direction is approximately 4m and the distance between cross-track scans is approximately 1m. The LMS-Q140i scan

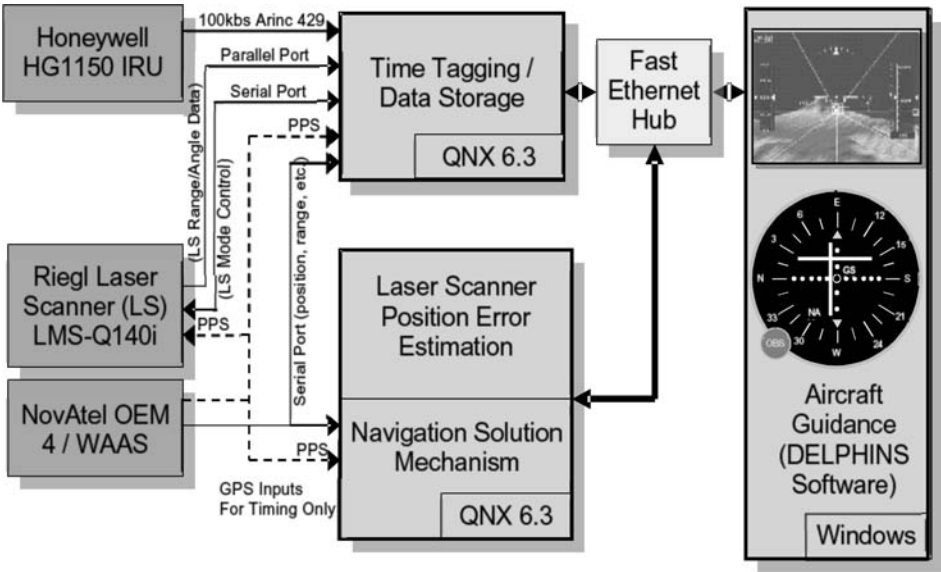


Figure 8.10 TERRAIN precision-approach hardware diagram.

rate can be set as high as 40 scans per seconds allowing for a more uniform distribution of the laser measurements. The specification for the LMS-Q140i can be found in [10].

Data from the various sensors were collected and time-tagged in the DCD computer. The DCD is configured to allow for data collection and distribution as well as postmission playback of the real-time TERRAIN guidance system. The DCD hardware consists of a 4U 19" rack-mountable computer configured to allow data collection from the NovAtel OEM 4/WAAS GPS receiver via an RS-232 serial port; a Honeywell HG1150 INS via an ARINC 429 data bus card; a Riegl LMS-Q140i laser scanner via an ECP parallel port and RS-232 serial port; and a Honeywell HG8505 radar altimeter via an RS-232 serial port. Along with data collection and distribution, another critical role of the DCD is data time-tagging. The DCD is synchronized to GPS time using the PPS signal from the OEM-4 GPS receiver and time information from the OEM-4 data messages. Time synchronization is performed by pulse-extending and level-converting the PPS signal to allow for the triggering of a serial port interrupt in the DCD. The constant length of the pulse extension is then corrected in the DCD software. Time accuracy in the computer is measured to be better than 10 μ s.

The aircraft state (position, velocity, attitude) and the guidance cues for an approach are computed in the NAV computer. The NAV computer receives sensor information from the DCD computer via the TCP/IP Ethernet link, processes and extrapolates the data, and transmits the aircraft state information to the display computer via a UDP Ethernet link. Like the DCD computer, the NAV computer is synchronized to GPS time via the OEM-4 GPS receiver's PPS signal.

Guidance cues are provided to the pilot through a liquid crystal display (LCD) mounted for use by the right seat pilot as shown in Figure 8.12. The LCD display is driven by DELPHINS software running on the tunnel computer. DELPHINS



Figure 8.11 DC-3 cockpit with DELPHINS tunnel display.



Figure 8.12 DC-3 on short final to runway 19, Braxton county airport (K48I).

software was developed and is maintained by Delft University of Technology in Delft, the Netherlands and used by Ohio University under a memorandum of agreement. It provides a heads-down-display (HDD) ,which can provide both a flight director style guidance display and a synthetic vision guidance display to the pilot.

In the development of the proof-of-concept TERRAIN precision approach system two goals were defined: the first was to operate in real time; the second was to provide aircraft guidance that meets the accuracy requirements necessary to conduct a precision approach. A precision approach is defined as providing both vertical and horizontal guidance. Aircraft approach systems are required to meet specific levels of accuracy, integrity, continuity and availability. The parameter levels are dependent on the category (CAT) of the approach. The CATs are specified by the height at which the pilot must visually acquire the runway environment and make a decision as to whether or not to continue the approach. This height is referred to as the decision height (DH), and the pilot's decision will be based on visibility and/or the runway visual range (RVR). For example, a CAT I approach has a DH of not lower than 200 ft and a visibility of not less a half statute mile or an RVR of not less then 1,800 ft. Thus, if the pilot cannot identify the runway environment at a height of 200 ft above the runway, or the ground controllers have measured the visibility at less than 1,800 ft, the pilot may not land using the CAT I approach. More information on aviation-related navigation, including WAAS, can be found in Chapters 4 and 10 of this book.

For the proof-of-concept TERRAIN precision approach system, the system's positioning accuracy was evaluated, using kinematic GPS positions as the truth reference. Figure 8.13 shows a set of position accuracies from one of the approaches.

It can be seen that for this approach the position estimates were nearly all contained within 6.5m of the truth. The shaded regions indicate when GPS WAAS

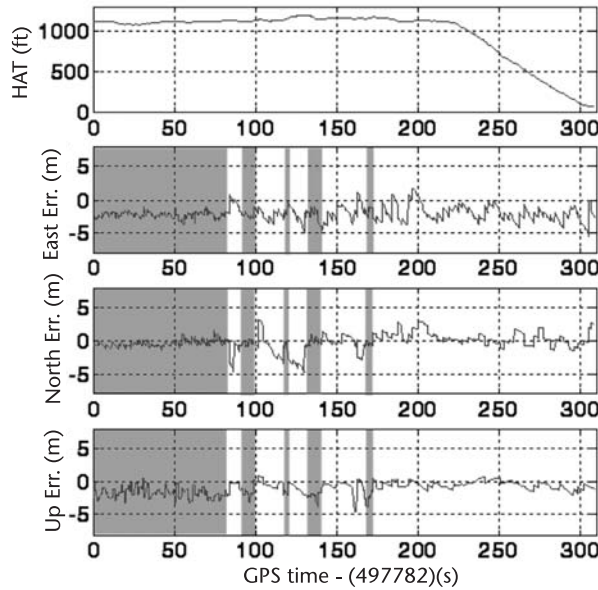


Figure 8.13 TERRAIN position—KGPS for one approach [height above threshold (HAT)].

positioning was enabled whereas the nonshaded regions indicate when TERRAIN mode was enabled. Mode switching between WAAS and TERRAIN can be observed. This effect can be attributed to variations in the terrain that cause the height AGL to exceed the threshold height for the WAAS mode, selected to be 1,200 ft HAT, at various occasions. It must be noted that the runway threshold is higher than much of the terrain on the approach path. The mode-toggling phenomenon results from placing the decision HAT for WAAS mode near the maximum range of the LADAR and would not occur with either a lower HAT or longer range LADAR. Table 8.3 presents a summary of the statistics from the eight approaches (nine minutes of data) flown on January 14, 2005, at Braxton County airport while the navigator was in TERRAIN mode. Values in the table were computed from data collected while on approach, from 900 ft HAT until the 50-ft DH.

Accuracy requirements defined for a satellite-based CAT IIIB approach are 5m in the horizontal and 2.9m in the vertical 95% value under guidance to the DH. If the data used to generate the standard deviations in Table 8.3 is assumed to be normal, then the 95% values can be conservatively approximated by multiplying

Table 8.3 Summary of TERRAIN Position Accuracy at 50-ft DH, Eight Approaches

<i>Error Direction</i>	<i>Mean (meters)</i>	<i>Standard Deviation (meters)</i>	<i>95% Value (meters)</i>	<i>Maximum Absolute Error (meters)</i>
East	-0.962	1.50	3.58	3.23
North	-0.031	1.99	3.60	2.82
Up	-0.685	1.33	3.35	2.16

the standard deviation by two and adding the absolute value of the mean error. Table 8.3's 95% values of the proof-of-concept TERRAIN precision-approach system meet the CAT IIb horizontal accuracy requirements, but it should be noted that this table represents only a small set of measurements (eight approaches) recorded to one runway. While the CAT IIb vertical accuracy requirement was not met, it is likely that errors in the INS-LADAR lever-arm translation and orientation estimates may have contributed to the error vertical bias given that precise lever-arm measurements were not available for the flight test. It is further noted that vertical accuracies can be improved by implementing techniques that use the runway profile after the threshold is crossed allowing centimeter-level vertical performance consistent with the laser scanner ranging accuracy specifications.

References

- [1] Baltsavias, E. P., "Airborne Laser Scanning: Basic Relations and Formulas," *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54, 1999.
- [2] Vadlamani, A., M. Smearcheck, and M. Uijt de Haag, "Preliminary Design and Analysis of a LIDAR-Based Obstacle Detection System," *Proceedings of the 24th IEEE/AIAA Digital Avionics Systems Conference (DASC)*, Washington, D.C., November 2005.
- [3] Smearcheck, M., A. Vadlamani, and M. Uijt de Haag, "Detection of Mobile Runway Obstacles Using Dual Airborne Laser Scanners," *Proceedings of the 27th IEEE/AIAA 27th Digital Avionics Systems Conference (DASC)*, St. Paul, MN, 26–30 October 2008.
- [4] Campbell, J. L., et al, "Light Detection and Ranging-Based Terrain Navigation—A Concept Exploration," *Proceedings from the 2003 ION GNSS*, Portland, OR, September 2003.
- [5] Campbell, J. L., M. Uijt de Haag, F. van Graas, "Terrain Referenced Precision Approach Guidance," *Proceedings of the Institute of Navigation's National Technical Meeting*, San Diego, CA, January 24–26, 2005.
- [6] RTCA Special Committee 193, "Requirements for Terrain and Obstacle Data (RTCA DO-276)," Washington, D.C., March 5, 2002.
- [7] Titterton, D. H., and J. L. Weston, *Strapdown Inertial Navigation Technology*, Peter Perigrinus Ltd. on behalf of the Institution of Electrical Engineers, 1997.
- [8] Hodgson, M. E., "Comparison of Angles from Surface Slope/Aspect Algorithms," *Cartography and Geographic Information Systems*, Vol. 25, No. 3, 1998, pp. 173–85.
- [9] Soloviev, A., F. van Graas, and S. Gunawardena, "Implementation of Deeply Integrated GPS/Low-Cost IMU for Reacquisition and Tracking of Low CNR GPS Signals," *Proceedings from the 2004 ION TM*, San Diego, CA, January 2004.
- [10] "Laser Mirror Scanner LMS-Q140-60-HR, Technical Documentation and User's Instructions," Edition 00-06, Rev 99-06-30, Riegl Laser Measurement Systems, Austria, 2001.

Combining GNSS with RF Systems

Richard Fuller

This chapter discusses the use of various radio frequency (RF) signals alongside GNSS to form integrated navigators. The chapter is written for the working engineer or student solving problems that involve finding the location of some article or person. A GNSS-based receiver in many cases is the preferred solution, but it may have some drawbacks in implementation. Possibly the precision needs to be higher than a stand-alone GNSS receiver, or maybe the application involves locating an article inside a building where GNSS signals cannot penetrate. For these reasons, researchers and developers have proposed a myriad of alternatives and augmentations to satellite-based location technology. Even the most exciting systems, however, have drawbacks when considering such things as global coverage, economical efficiency, and technological stability next to the overall accuracy of GNSS technology.

As will be discussed in the next section, there is no single alternative that represents a true “replacement” for GNSS location capability, but there are many that complement and augment satellite-based systems well. This chapter provides examples of the different types of approaches that can be combined with GNSS to create a system meeting requirements that are not achievable by GNSS alone. First, the chapter outlines several approaches that use RFs to either estimate location or help a connected GNSS determine position. Next, the chapter discusses some of the mathematical techniques used to combine GNSS with other RF systems. Finally, the chapter presents a number of specific example systems that combine GNSS with other RF location technologies. Appendix 9A discusses the propagation characteristics that need to be taken into account when designing a system that uses RF signals for location determination.

9.1 Location System Alternatives

When considering the alternatives for augmenting a GNSS receiver’s capability, the primary question that must be answered is: “Why use RF systems with GNSS versus other non-RF approaches?” First, RF represents some of the most flexible kinds of systems that can be deployed versus the alternatives. For example, ultrasound-based systems can have a very fine precision but are strictly limited to LOS operation and highly subject to environmental noise [1]. Table 9.1 summarizes the benefits and drawbacks of different non-RF systems deployed today.

Table 9.1 Comparison of Different Location Technologies [1, 2]

<i>Technology</i>	<i>Pros</i>	<i>Cons</i>
Inertial	<ul style="list-style-type: none"> — Precise with expensive sensors — Works in a variety of environments (underwater, indoor, etc.) 	<ul style="list-style-type: none"> — Unusable without frequent corrections from external reference (often GNSS-based)
Vision	<ul style="list-style-type: none"> — High-precision — Works where instrumented 	<ul style="list-style-type: none"> — Limited range, LOS — Extensive instrumentation
Ultrasound	<ul style="list-style-type: none"> — Very inexpensive emitters and sensors — High-precision 	<ul style="list-style-type: none"> — Limited range, LOS — Very sensitive to ambient noise
Infrared	<ul style="list-style-type: none"> — Inexpensive emitters and sensors — High-precision 	<ul style="list-style-type: none"> — Limited range, LOS — Very sensitive to ambient noise
Smart floors/furniture	<ul style="list-style-type: none"> — Very precise — Works without instrumenting subject 	<ul style="list-style-type: none"> — Works only where floor (or furniture) is instrumented — Difficult to distinguish individuals (i.e., track a specific person)
GNSS	<ul style="list-style-type: none"> — Worldwide coverage — Free to use 	<ul style="list-style-type: none"> — Precision 10m unaided — Limited indoor use

In most cases the most limiting factor is the transmission range of the signals. Since RF propagation can be used for close-in applications as well as those applied over hundreds of kilometers, it is the most flexible method used in location systems. However, in many cases where RF has superior range it does so at the expense of precision, which is why alternative means of location determination will always be sought. One such example in wide use today is the complementary blending of inertial sensor and GNSS measurements. This combination can offer very precise location capability without the need for installed infrastructure. This is limited by the precision of the GNSS system it is paired with as well as the availability of signals to calibrate the inertial sensors. The technique of combining GNSS with inertial sensors is discussed in detail in Chapters 6 and 7.

Other aspects of a RF-based location system that must be considered include transmission frequency and associated antenna size, multipath and signal penetration into buildings or other structures. A detailed discussion is provided in Appendix 9A (included on the DVD), while a summary of these issues are presented in Table 9.2.

Frequencies at MF and below have beneficial signal propagation performance, including good building penetration. However, since efficient antennas at these frequencies are very challenging to design, high transmit power is required to achieve the required signal level at the receiver [3]. For example, the Loran system (discussed in greater detail in Chapter 11) in the LF band at 90–110 kHz with a transmit power at the stations of 1 MW [1]. While this high power is necessary to operate over long distances (typically 900–1500 km over land, 1500–2000 km over sea [4]) it is also to compensate for antenna losses that could easily exceed 60 dB for a handheld device or 30 dB for a fixed 1-m receiver [3].

While at lower frequencies, efficient antennas are key; at higher frequencies, the local environment becomes the driving factor in location determination. Above

Table 9.2 Characterization of RF Location Systems in Various Frequency Bands

<i>Band</i>	<i>Frequency</i>	<i>Propagation</i>	<i>Antenna Size (Half-Wave Dipole)</i>	<i>Multipath</i>	<i>Structure Penetration</i>
VLF	<30 kHz	LOS, surface, sky	5 km	Ground, sky	Good
LF	30–300 kHz	LOS, surface, sky	0.5–5 km	Ground, sky	Good
MF	300 kHz–3 MHz	LOS, surface, sky	50–500m	Ground, sky	Good
HF	3–30 MHz	LOS, surface	5–50m	Ground, artifacts	Poor
VHF	30–300 MHz	LOS	0.5–5m	Artifacts	Poor
UHF	300 MHz–3 GHz	LOS	5–50 cm	Artifacts	Very poor
SHF	3–30 GHz	LOS	0.5–5 cm	Artifacts	Low/none

100 MHz, LOS signals become the dominant propagation type where buildings, vehicles and natural obstructions like trees begin to have greater effect [5]. Multipath becomes a major concern as reflections from these obstacles arrive both in-phase and out-of-phase at the receiver causing fading noise. Above 3 GHz there is only the option of direct LOS propagation since nearly all obstacles will highly attenuate the signal.

9.2 RF Location Types and Classifications

Numerous types of RF location technologies have been implemented over the years. As illustrated in Figure 9.1, these primarily break down into five major types: (1) proximity; (2) direction-finding (DF) or angle of arrival (AOA); (3) Doppler; (4) signal strength; and (5) timing or phase.

Proximity-based approaches include contact and near-contact sensors such as in RF identification (RFID). In the case of RFID, these systems establish location based purely by presence. These systems usually operate in the near field within a few wavelengths. If the tracked transmitter object is close enough for a receiver to get a signal, then the receiver and transmitter are clearly close to one another.

DF and AOA systems provide a means for determining the bearing of a RF transmitter from a receiver. Two or more receivers can be used to triangulate on the two-dimensional horizontal location of the transmitter. Additionally, altitude could be determined by a third, vertically oriented receiver.

Location by Doppler uses the phenomenon that when a transmitter and receiver are traveling towards one another the received frequency is higher than transmitted, and when they are traveling away from each other the received frequency is lower. By combining multiple measurements of frequency shifts it is possible to ascertain the location and velocity simultaneously.

Signal strength uses signal power or other signal-based metrics like bit error rate (BER), or carrier-to-noise ratio (CNR) to form an estimate of range, often referenced to previously stored location information in a database, otherwise known as *RF fingerprinting*. In the ranging case, multiple range estimates can be combined via lateration from multiple reference positions. In the case of RF fingerprinting, previously collected reference values are compared to one or more current measurements to match the current location to the stored data.

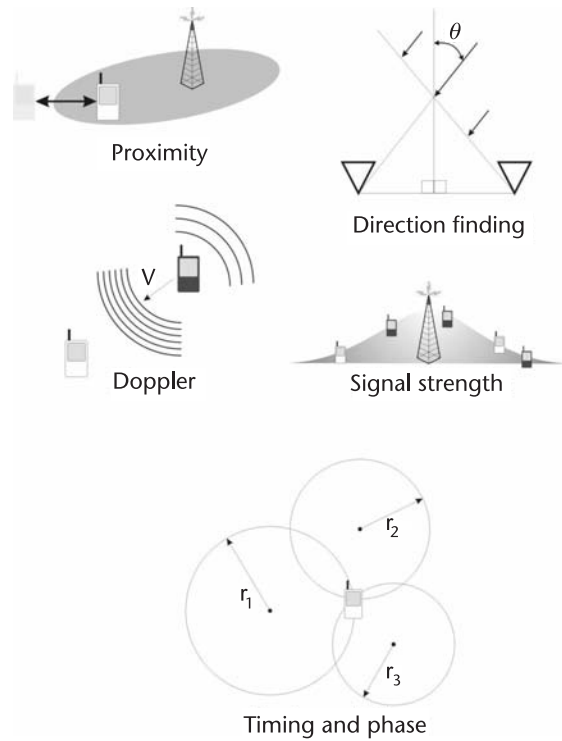


Figure 9.1 Overview of the different types of RF location technologies.

Timing and phase location systems use measurements of the received phase of an RF signal or an additional timing modulation on the signals to estimate the range between a transmitter and receiver (as is the case for GNSS). Lateration is used to combine multiple range estimates to form a location estimate. This timing information can be employed directly as in time-of-arrival (TOA) systems, or in round-trip-delay (RTD) measurement, or alternatively in time difference-of-arrival (TDOA) systems.

The following five sections will discuss in further detail how these different approaches work and look at them in light of these characteristics:

1. Applicable range;
2. Precision;
3. Types and sensitivity to interference;
4. Passive versus active utilization;
5. System capacity and scaling;
6. Forward versus reverse links.

9.2.1 Location by Proximity

One of the simplest approaches to RF location is simply by sensing a RF beacon transmitter that is at a known location. The location precision of the receiver is the coverage area of the beacon transmitter. Depending on the propagation

characteristic of the frequency, channel characteristics, directionality of the transmitter/receiver pair, and the effective aperture of the transmitting and receiving antenna, the range may be extend from a few meters to hundreds of kilometers.

The type of device most widely in use today representing proximity-based location is RFID as depicted in Figure 9.2. An RFID system is comprised of at least one *reader*, which operates both as a power transmitter and receiver, and one or more passive *tags*, which are used as transmitters [1]. The system takes advantage of the previously described attributes of near-field RF patterns that allow for electrical and magnetic coupling of the transmitter and receiver. The RFID tags are not usually powered by batteries or other power sources but derive their power from the reader RF signal. This allows the tag to power a small microchip-based circuit capable of modulating a unique ID code or some other stored information unique to the tag. Depending on the power transmission of the reader the range to activate the tag can be as little as a centimeter or as large as 30m or more [6]. An alternative to passive tags where the reader provides the power to the tag are active RFID tags that have their own power and broadcast information at periodic intervals or when polled by the reader. This approximates a more traditional telemetry system, with the exception that the tags are designed for minimal complexity, storage, and data bandwidth in order to keep their production costs down. RFID operates in different frequency bands; in low-frequency operations it is employed at 30–500 kHz and is used mainly for short-range reads (less than 1m), while at higher frequencies it can operate in the range of 850–950 MHz and 2.4–2.5 GHz; such operation is better suited for longer read ranges.

A novel application that uses a combination of GPS and RFID is the YARD HOUND by PINC Solutions [7], which is in commercial use today for management of tractor trailers at distribution centers as shown in Figure 9.3. This system places passive RFID tags on tractor trailers either permanently or when they enter the yard of a distribution center. A reader is placed on each of the yard trucks that move the trailers between storage bays and the distribution warehouse as well as the exit and entrance to the yard. The yard trucks also have precise GPS location so that when the reader records contact with a particular tag, it can estimate the location of a trailer to roughly one storage bay, thus creating a record of the current location of each trailer that it passes. These records are transmitted via a 802.11 b/g/n

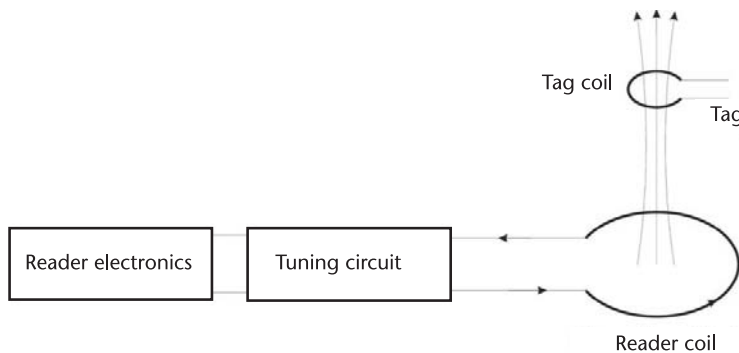


Figure 9.2 Overview of RFID configuration.

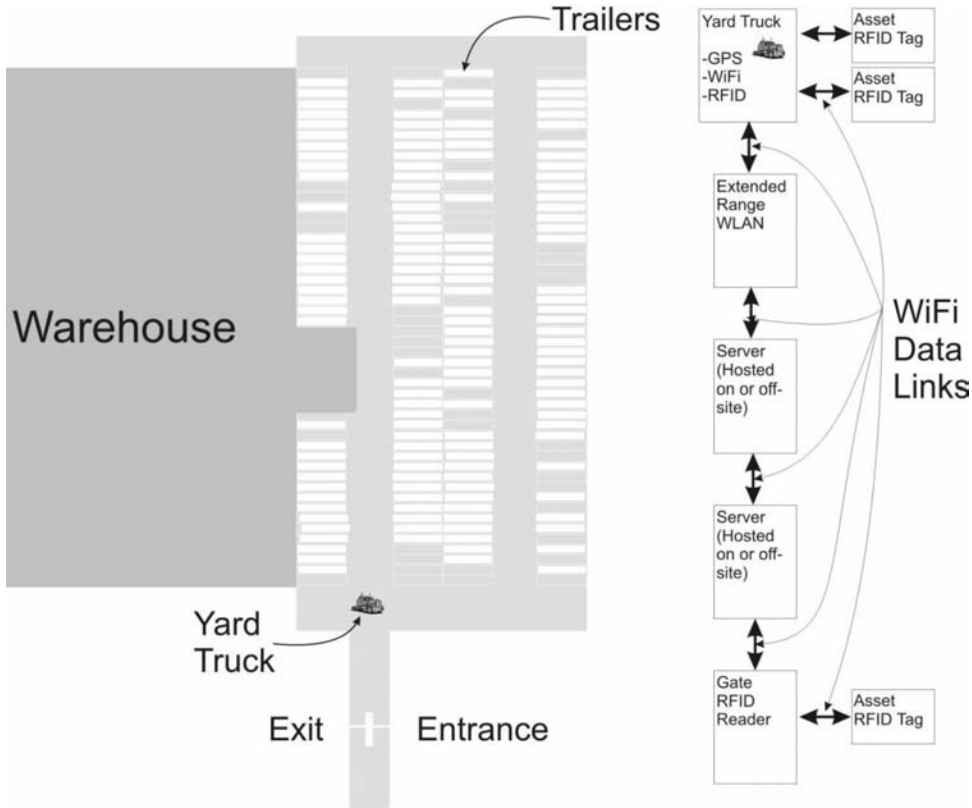


Figure 9.3 Illustration of the YARD HOUND™ system.

wireless link from the yard trucks to a monitoring computer (running in the distribution warehouse), which hosts a Web-based interface that allows both managers inside the warehouse sitting at computer terminals as well as workers with mobile devices to get instant access to the location of the trailers they need to either fill or empty. This dramatically decreases the time and effort of personnel to survey the yard to map the location of trailers at a given epoch.

Bluetooth (IEEE 802.15) is another wireless technology for short-range communications. It uses a frequency hopping scheme that makes timing/phase measurements very difficult, but it is well-suited for proximity-based sensing since its range is dependent on its power class (optionally 1, 10, or 100m) [8, 9]. Bluetooth operates in the ISM band between 2.400 GHz and 2.485 GHz so it is substantially, but not completely, degraded when penetrating structures. Each Bluetooth device is identified by a unique media access control (MAC) address, allowing accurate identification of objects. The advantage of Bluetooth is that it is a pervasive technology in consumer and commercial applications. Many of today's cell phones, computers, printers, and other electronic devices come pre-equipped with Bluetooth. The most obvious combination of this technology with GNSS is the use of this system as a complementary location system for use indoors. The device could be designed to combine the best available data at the current time. For example, if a

precise GNSS position estimate were available it would be used. Alternatively, if a GNSS solution were unavailable or unreliable, the Bluetooth radio would seek out MAC addresses and compare them to a database of known locations to estimate the user location. Similar to Bluetooth are WiFi (802.11b/g/n) and ZigBee (802.15.4) RF signals, which also operate in the ISM band of 2.400–2.485 GHz, and can also be used as proximity sensors in combination with GNSS [10]. Section 9.2.4 discusses further applications for all three of these RF radio systems.

In general, proximity-based location can be classified as an ultra-local or local area system where the range generally would not exceed 100m. Precision of the location estimate would be inversely proportional to range. Considerations of bandwidth, signal attenuation in the atmosphere, or through structures are generally not important. This is because the devices are often operating in the near-field of the RF bands, making far-field effects inconsequential. Since the relative distance of the receiver and transmitter is reduced, antenna gain is usually not a substantial issue, thus permitting antennas to be small and low-power. Of particular note, relative to GNSS, is that these systems always put a constraint on the maximum number of simultaneous users for a fixed infrastructure, since they all depend on some external reader or base station with a limited number of channels. Correspondingly, these systems increase in size and complexity with an increase in simultaneous users.

9.2.2 Location by Radio Direction Finding (DF) and Angle of Arrival (AOA)

The basis of radio DF and AOA position determination is the use of angular measurements for triangulation. Triangulation uses the bearings from two or more DF receivers to obtain an estimate of the transmitter's location. The geometry of such a system is illustrated in Figure 9.4. These systems principally operate in the RF far-field of the transmitter with a planar wave incident on the receiver. In practice, multiple signal distortions prevent this configuration from being ideal.

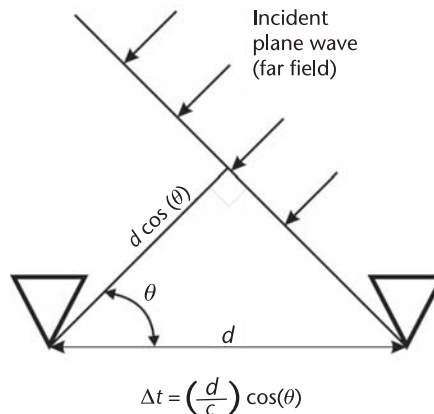


Figure 9.4 Radio DF and AOA principle.

Work in radio DF dates back over 100 years. The earliest recorded systematic experiments were in 1899 with the first demonstration of tracking a steamship in 1906 by Marconi using a radial system of horizontal antennas about one-fifth a wavelength long. In 1907, work by Pickard and DeForest showed the benefits of loop antenna direction finders, which proved to be the format of choice for many small-aperture DF systems (at antenna sizes up to about half a wavelength). In 1926, the Watson-Watt DF was introduced. This used two crossed, quadrature loops with their output signal directly measured by a cathode-ray oscillograph providing an instantaneous estimate of AOA [5].

In the 1930s leading up to World War II, medium- and high-frequency radio beacons served as guidance and rough navigation functions for aircraft. By tracking a single beacon (at an airport for example) a relative angular fix can be made between the axis of the aircraft and the beacon. By tracking two or more beacons, a crude position was determined through triangulation. “Positive” navigation points were provided at some airports with vertical marker beacons. These vertical beacons were similar to the standard radio beacons except that their signals were directed upward. Tracking this beacon indicated to pilots that they were directly above it, thus giving them “positive” feedback on their location. As more flights were made on a greater number of routes, the use of airport locations as beacons became undesirable. Over time, directional beacons were set up along air travel routes around the world. These routes took advantage of the lessons learned during World War II, during which it was shown that routes directed along prevailing winds (referred to as cyclonic) could cut travel time over the shortest distant method (great circle), thus saving fuel. After World War II, VHF omnidirectional range (VOR) stations were developed and put into place. According to [11], outside of the developed world VOR support is provided with the exceptions of the polar regions, the south Atlantic Ocean, and much of the Pacific and Indian oceans.

VOR broadcasts in a band between 108 and 118 MHz with channels 50-kHz-apart. The ground broadcasts two signals modulated with a 30-Hz tone. The first is a fixed reference tone with an omnidirectional radiation pattern. The second signal is radiated with a cardioid pattern that rotates 30 rotations per second. The receiver uses the relative phase of these two 30-Hz signals to determine its bearing towards the VOR (which is at a known location). Figure 9.5 shows a diagram of a VOR receiver [12]. In addition to the 30-Hz modulation there is an audio channel broadcast modulated at 9,660 Hz as well as a Morse code identifier of the VOR with a 1,020-Hz modulation. The voice signal can be used as a repeating station identifier or as a ground-to-air communications channel [4]. This system design was a U.S. standard by 1946 with later adoption by the International Civil Aviation Organization (ICAO). The performance of this system is only limited by propagation effects and user equipment errors. High-end user equipment can usually achieve 0.1° – 10° of angular resolution. This system only works well when the VOR and receiver are LOS-visible [11].

A unique combination of GNSS and VOR technologies was proposed in [13]. The combination used a GPS receiver connected to a VOR signal generator to create a mobile short-range VOR broadcast terminal. As shown in Figure 9.6, the goal of the system is to create a signal that is input to an aircraft’s VOR system, which has position, velocity, heading and cross-track information from the GNSS

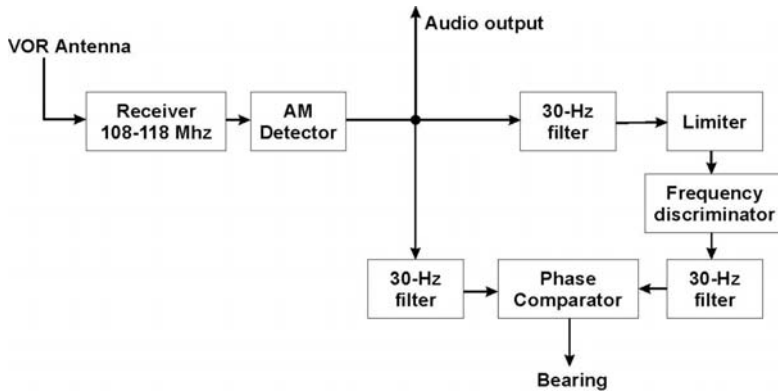


Figure 9.5 Diagram of a VOR bearing estimation receiver.

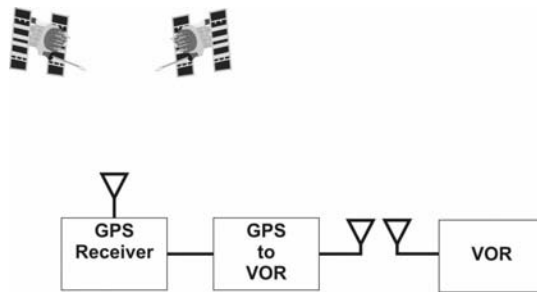


Figure 9.6 GPS to VOR concept.

receiver. To the pilot, the use of the VOR instrumentation is the same as all-terrestrial signal sources. While not explicitly stated, this system design suggests a more general AOA beacon system that could be deployed in operational scenarios. This system would allow two or more mobile RF beacons with a location reference tied to GNSS to be deployed to enable one or more AOA receivers to operate locally. Of course, such a system would have to overcome the serious limitations to DF/AOA usage, primarily multipath-shifting the received angle of the signals and usually would require LOS operation.

One positive advantage of the DF/AOA system is that most of the architectures are designed with no limit on the number of users within an operating region. It is interesting to note that the recent advent of advanced antenna and RF front-end technologies in WiFi, most notably multiple input multiple output (MIMO) systems with two or more antennas, have led to an interest in AOA for consumer and commercial equipment [1].

9.2.3 Location Using Doppler Frequency

Another method that can be used to locate a RF device utilizes the well-known phenomenon called the Doppler effect. The Doppler effect states that when an object emits a frequency and moves relative to an observer, the frequency of the

observed signal will be shifted up or down. The magnitude and sign of the shift depends on the frequency of the signal and the velocity of the transmitter and observer relative to each other. The Doppler shift is given by:

$$\Delta f = (f_R - f_T) = -\frac{v}{\lambda} \quad (\text{Hz}) \quad (9.1)$$

where f_R is the received frequency, f_T the transmitted frequency, λ is the signal wavelength and v is the relative velocity between the transmitter and receiver.

As illustrated in Figure 9.7, the velocity of a transmitter at a given epoch, \mathbf{v}_1 , is projected onto the LOS, $\hat{\mathbf{r}}_1$, to a receiver at (x_p, y_p) that produces a measured Doppler shift, $\text{dot}(\mathbf{v}_1, \hat{\mathbf{r}}_1)$. As the location and velocity of the transmitter and the Doppler shift are known, the only unknown is the location of the fixed receiver. At a later epoch another Doppler shift measurement is made from a different transmitter location, and these two observations can be used to resolve a two-dimensional location. Note that this assumes perfect frequency references for both the transmitter and receiver, a stationary receiver, and no frequency distortion between transmitter and receiver. Additionally, when dealing with a transmitter and receiver at large distances from one another (as in GNSS) the LOS is not very sensitive to the location of the receiver so precise position is hard to estimate from a single epoch of Doppler data [14]. For a short-range Doppler system, where the distance between receivers and transmitters is within an order of magnitude of the desired location precision, Doppler location is far more precise.

Doppler positioning was the basis of the first navigation satellite system deployed, called Transit [15, 16], which was decommissioned by the U.S. Navy at the end of 1996.

Another interesting approach presented in [17] utilizes a mobile node that transmits a signal at a known frequency. Here, the Doppler-shifted signal is

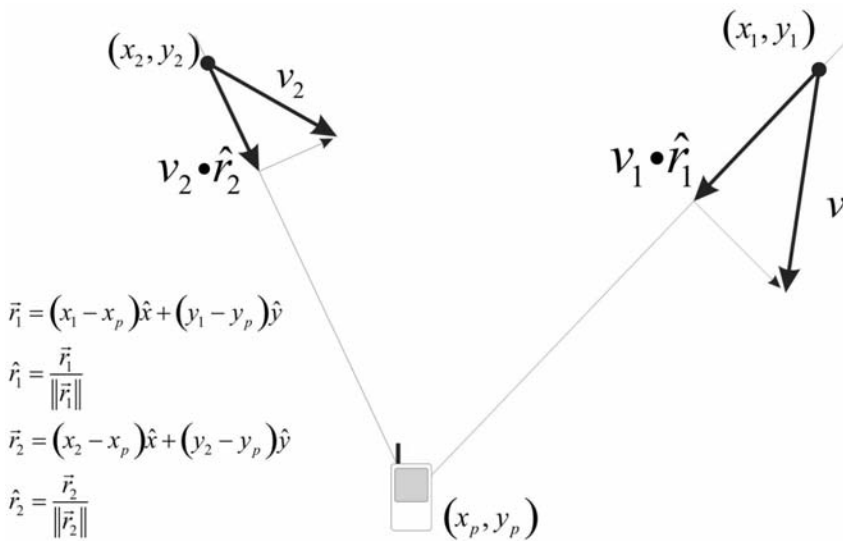


Figure 9.7 Doppler location estimation using measurements at different times.

measured by additional fixed “infrastructure” nodes. The speed of the mobile node relative to all infrastructure nodes can be calculated and used to determine both the velocity and the location of the mobile node. The desired hardware in this approach had low computational power and no specialized Doppler measurement capability. The approach proposed in [17] is depicted in Figure 9.8; the tracked node transmits at a frequency f_t and one of the infrastructure nodes generates a signal f_a such that $f_t > f_a$. Further, when the two signals interfere with one another they produce a signal with an envelope frequency of $f_t - f_a$. This frequency difference can be measured with very fine precision using simple direct sampling at sensor nodes S_1 , S_2 , and S_3 .

Results of simulation and experiment showed results of 1.3–2.2m and 0.1–0.4m/s, respectively, in an operational area of 1,500 m². The limitation of this approach is that each tracked node needs to use a specific frequency (or share it in a synchronized time-sliced method) and thus cannot support an unlimited number of nodes (in contrast to a direct-Doppler system like Transit, which supported an unlimited number of tracking receivers).

The advantage to the Doppler systems is that they are relatively simple in overall design. Doppler-based location systems can be deployed in a worldwide coverage region, which was the case for Transit, or in local-area contexts such as a 50m x 30m (1,500-m²) warehouse. LOS requirements are the same as for any RF-based system; however, the frequency choice is fairly flexible since it is not limited to a particular modulation or bandwidth.

9.2.4 Location Estimation Using Signal Strength

A substantial body of work has been produced examining position-determination techniques based on signal strength and other signal-related properties such as signal-to-noise ratio (SNR), bit-error-rate (BER), link quality indicator (LQI), response rate (RR), or carrier to noise ratio (CNR). The two most prevalent RF sources used in these studies are WiFi access points (IEEE 802.11) and cellular towers (both GSM and CDMA).

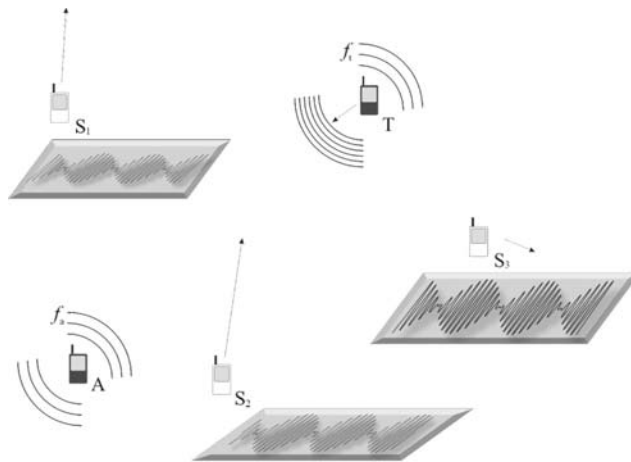


Figure 9.8 Interference-based Doppler velocity and location.

There are basically two general classes that signal strength solutions break down into: distance estimation (leading to lateration) and pattern recognition (also known as RF fingerprinting).

9.2.4.1 Distance Estimation

As the distance between the transmitter and the receiver increases, the received signal strength (RSS) decreases as discussed in Appendix 9A.2 and in more detail in [18, 19]. Figure 9.9 illustrates the decreasing and fluctuating signal level (due to both the environment and fading) as it travels from the transmitter. Using this profile, the RSS can be linked to a distance measurement. In cellular and WiFi systems the RSS is used for handoff and traffic processing and is accessible without changing system architecture. Operating using RSS is subject to severe local fading in urban areas and is usually unreliable in many instances [19]. The path loss of the signal strength is dominated by free-space path loss over large distances (generally hundreds of signal wavelengths) but can observe large variations in tens of decibels over shorter lengths (typically less than a few wavelengths) due to short-term fading. Long-term fading (from a few wavelengths to over 100) can experience fading due to reflections in the local environment such as buildings and refraction of indirect paths in the atmosphere [20]. A specific example of a system developed using signal strength is the RADAR system developed by Microsoft Research that used 802.11 access points (AP) to estimate the location of a mobile transmitter within an office environment [21]. Initially, signal strength measurement data was collected as a function of location throughout one floor of an office building to build a model during the “training phase” that could subsequently be used to estimate location. When discussing indoor location systems, it is necessary to model the transmission properties through walls and other obstructions [22]. Additionally, the relative orientation of the receiving antenna and transmitter can greatly affect the received signal strength. By using a model that was generated using orientation data as well, they were able to show location errors around 2 to 3m for the office environment, in many cases good enough to isolate the mobile transmitter to a specific room. The drawback of this approach is the training phase (including retraining if the

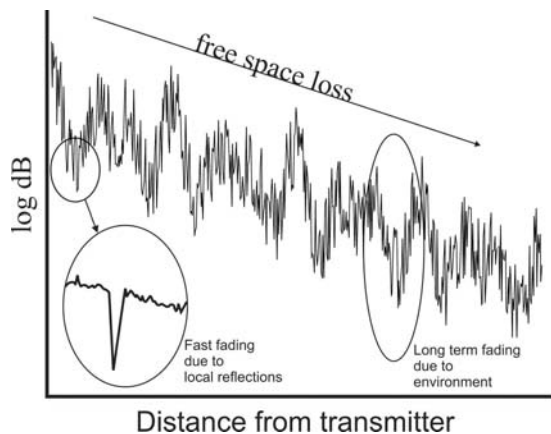


Figure 9.9 Propagation path loss with multipath.

physical layout should change) and the variability of environmental factors such as the number of people in the office building at a given time [21].

Recent work in ray-tracing techniques have shown some promise to lower (or eliminate) the need for training in signal strength approaches as long as adequate data is available in the operational area. Notably, the overall accuracy compared to timing-based approaches is generally lower [23]. However, once the range is estimated, it can be used in the same fashion as timing-based methods. It is important to note that other measurements in addition to RSS can be used for range estimation; these include BER, LQ, and response rate (RR). For example, [24] discusses a method where the RR, which is the percentage of time that a given transmitter was heard in all scans at a specific distance from that transmitter, is used to estimate range for isolation of a Bluetooth device to a particular room on an office floor.

9.2.4.2 Pattern Recognition

In contrast to approaches that seek to map signal strength into range, pattern-recognition techniques have been applied to signal metrics to form location estimates. This type of approach is often referred to as *fingerprinting* and uses previously stored measurements or calculations to map locations for later matching to observed measurements. This approach takes advantage of the local-area fading characteristics by using features from multiple transmitters to uniquely identify a location. Generally speaking, the finer the training grid the better the resultant accuracy [23, 25]. Obviously, the first crucial step is formation of the radio map during a training phase, where either an empirical or model-based approach can be used. As described in [26], for an empirical approach the data can be collected as single values at a given location or accumulated as a probability distribution. Alternatively, model-based methods can categorize based on parameters. In a model-based approach parameters can either be provided a priori as in [21] or estimated from a small set of estimation fingerprints as in [27]. The propagation model can either be a direct-path only approach [21], or it can use multiple paths such as ray tracing [23]. Finally the representation can be based on single values [28] or in the form of probability distributions [29].

Pattern-matching solutions have been employed on a variety of scales, on the meter level within a single building [25] or over larger ranges with lower overall accuracies [30]. An important factor about path loss and pattern matching in signal-strength approaches is that they are not particularly sensitive to the frequency employed. There still are restrictions involving attenuation and LOS issues at different frequencies, but as long as the signal can be measured at the receiver, it can be used in a pattern recognition or signal strength-based location technique. Another strong suit of the pattern recognition approach is that it is not restrictive on the number of receivers that can make use of the signal measurements, thereby allowing for a large number of users for a given infrastructure.

9.2.5 Location Using Time, Phase, and Differential Timing of Arrival (TOA, POA, and TDOA)

GNSS systems use pseudoranges, timing measurements collected at a receiver (described in Chapter 3) that can be based on time-modulated spreading code or

the carrier phase. This is the most obvious example of a time-of-arrival (TOA) and phase-of-arrival (POA) system. However, such systems are not exclusively limited to satellite-based navigation systems. Typically multiple TOA ranges are combined via multilateration. The range measurements are represented for an individual transmitter by:

$$\rho^n(t) = r^n(t) + b(t) + \epsilon \quad (9.2)$$

where n is used to indicate the transmitter, which is not necessarily a satellite (denoted as k in the notation of Chapter 3). In navigation systems this is the well-known pseudorange, $\rho^n(t)$, between the receiver and the n th transmitter, as shown in Figure 9.10. The geometric distance between the transmitter and receiver is given by $r^n(t)$ and $b(t)$ is the combined clock offset in meters between the receiver and transmitter from a reference time (such as GPS time), often referred to as the *clock bias*. The additional error terms, as included in (3.1) have all been absorbed into the single variable ϵ for simplicity. When examining the three-dimensional location of the receiver, the geometric range can be represented in a similar manner (3.3), where the three-dimensional position of the receiver and transmitter k are represented as $\mathbf{x}_u = [x, y, z]$ and $\mathbf{x}^n = (x^n, y^n, z^n)$, respectively.

The clock bias is a combination of the receiver clock offset from a reference time as well as the n th transmitter's clock offset. A more detailed explanation of GNSS position, velocity, and time estimation using pseudoranges can be found in Chapter 3.

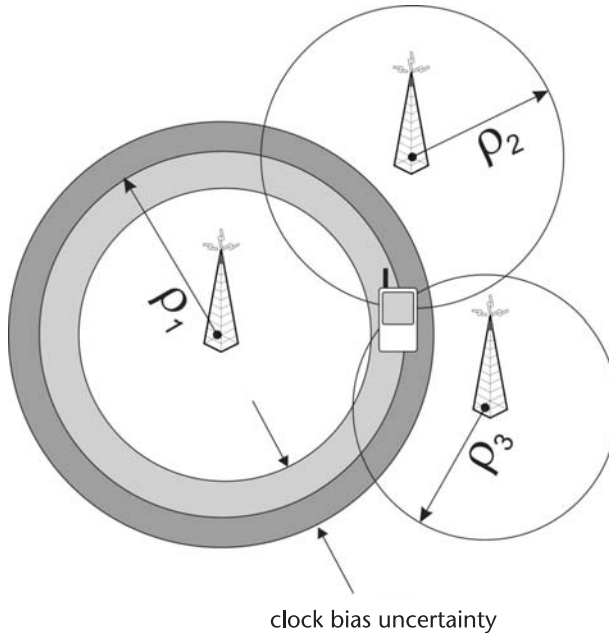


Figure 9.10 Location from TOA and POA ranges.

The same formulations as for TOA systems apply for POA systems; however, there is usually a carrier-cycle ambiguity between the transmitter and receiver as shown in Figure 9.11. The phase ambiguity that occurs in these systems is the same in principal as the GPS phase ambiguity when making a phase-based range estimate with GPS, as described in Chapter 14. However, one of the advantages of POA systems is the precision of the carrier phase. In GPS, for example, the L1 code phase cycle length is roughly 300 km in range, and the carrier wavelength is 19 cm. This enables far more precision in ranging measurements at the expense of difficulty in resolving integer ambiguities [14, 31]. For example, if there is an uncertainty of 1m in the initial position estimate there could be as many as 10 integers for a given satellite and 10^n combinations for n satellites, which could easily lead to millions of possible combinations [32].

In contrast to TOA, time-difference of arrival (TDOA) systems use arrival times from two transmitters at a receiver. Figure 9.12 shows that the time differences in a TDOA system as represented by line-of-position hyperbolic curves along which the receiver lies. By using a second measurement the two-dimensional position can

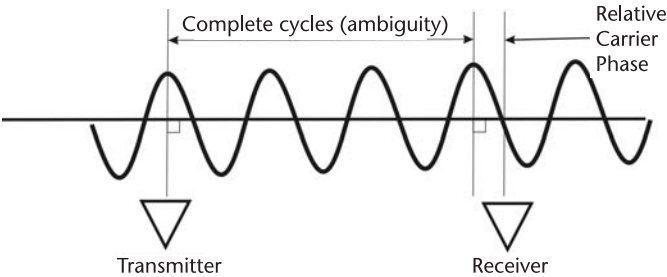


Figure 9.11 Illustration of the carrier-phase ambiguity between transmitter and receiver.

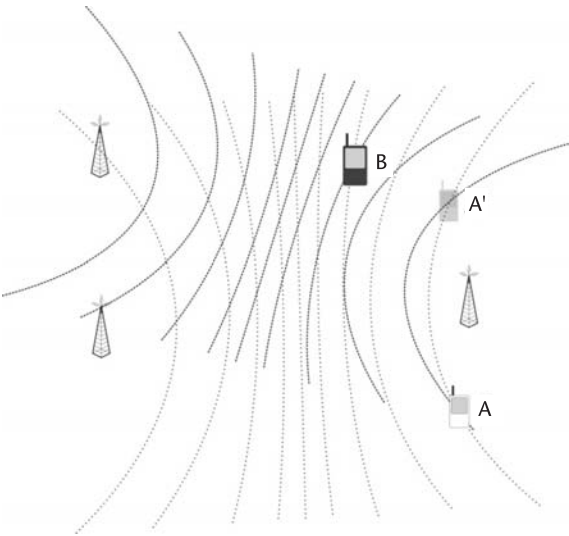


Figure 9.12 Location from TDOA hyperbolic traces.

be determined; this is shown in Figure 9.12 as receiver B. In the case of points A and A', a single difference measurement may not be sufficient to resolve the ambiguity, with respect to another difference measurement [4]. A significant advantage of TDOA over TOA is that the receiver clock bias is not an important factor in location determination, since as can be seen from (9.2), the receiver clock term $b(t)$ will cancel upon pseudorange differencing. (Note that the transmitters are referenced to a common source, such as GPS.)

A class of RF systems that has received a great deal of attention in the literature in recent years is ultra-wideband (UWB). UWBs are signals that possesses an absolute bandwidth greater than or equal to 500 MHz or a bandwidth exceeding 20% of the carrier frequency. UWB signal transmissions received limited approval by the U.S. Federal Communications Commission (FCC) in 2002 [33] for data communications, radar, and safety applications across a band from 3.1 GHz to 10.6 GHz. The power limits on the FCC order relegates current UWB applications in these bands to indoor and short-range applications. As is discussed in Appendix 9A.3, the high bandwidth of UWB systems allows extremely precise timing measurements to be made. The difficulty arises in implementing a TOA estimate using low sampling rates (versus the bandwidth). The suggested reading appendix included on the DVD includes a section on UWB Location Systems for those interested in various methods of determining TOA measurements from UWB signals.

9.3 Estimation Methods

Throughout the following section, the following convention is followed: Estimation includes approaches where location information can be broken down into value-based deterministic solutions and probabilistic random-process methods. Integration involves combining different location measurements or estimates into a combined solution.

9.3.1 Deterministic Estimation Using Triangulation

The term *triangulation* refers to two different location approaches. The first is lateration, involving ranging (pseudorange for the case where systematic errors exist besides measurement errors) from known points (TOA, POA, TDOA). The second is angulation, which measures angles or bearing relative between points (AOA and DF).

9.3.1.1 Lateration

For lateration, ranges can be combined from multiple systems as long as timing is referenced back to a common master clock (such as GPS time). In this instance range measurements from sources other than GNSS satellites can be combined into the position estimation steps to estimate the receiver location. See Chapter 3 for a detailed explanation of this procedure.

9.3.1.2 Angulation

In a likewise manner for angulation in two dimensions,

$$\theta_c^n = \tan^{-1} \left(\frac{(y^n - y)}{(x^n - x)} \right) + \tilde{\epsilon}_\theta^n \quad (9.3)$$

uses a nonlinear equation solving for position (x, y) from $n = 1$ to N angular observations θ_c^n , for the n th transmitter, as shown in Figure 9.4. Using a linearization method around a nominal location, (x_0, y_0) , we can represent the angle measurement as,

$$\theta_0^n = \tan^{-1} \left(\frac{(y^n - y_0 - \Delta y)}{(x^n - x_0 - \Delta x)} \right) \quad (9.4)$$

where

$$\begin{aligned} x &= x_0 + \Delta x \\ y &= y_0 + \Delta y \end{aligned} \quad (9.5)$$

with $(\Delta x, \Delta y)$ now representing the correction to apply to the nominal location. If this is expanded using a Taylor series and only keeping the linear terms it leads to,

$$\begin{aligned} f(x, y, z) &\equiv f(x_0 + \Delta x, y_0 + \Delta y) \\ &= f(x_0, y_0) + \frac{\partial f(x_0, y_0)}{\partial x_0} \Delta x + \frac{\partial f(x_0, y_0)}{\partial y_0} \Delta y + \dots \end{aligned} \quad (9.6)$$

where the partial derivatives are given by [34]

$$\begin{aligned} \frac{\partial f(x_0, y_0)}{\partial x_0} &= -\frac{y^n - y_0}{(r_0^n)^2} \\ \frac{\partial f(x_0, y_0)}{\partial y_0} &= \frac{x^n - x_0}{(r_0^n)^2} \\ r_0^n &= \sqrt{(x^n - x_0 - \Delta x)^2 + (y^n - y_0 - \Delta y)^2} \end{aligned} \quad (9.7)$$

Differencing (9.4) and (9.3) and combining with (9.6) and (9.7) results in,

$$\begin{aligned} \delta\theta^n &= \theta_c^n - \theta_0^n \\ &= f(x_0, y_0, z_0) + \frac{\partial f(x_0, y_0, z_0)}{\partial x_0} \Delta x + \frac{\partial f(x_0, y_0, z_0)}{\partial y_0} \Delta y + \tilde{\epsilon}_\theta^n \\ &= -\frac{y^n - y_0}{(r_0^n)^2} \Delta x + \frac{x^n - x_0}{(r_0^n)^2} \Delta y + \tilde{\epsilon}_\theta^n \end{aligned} \quad (9.8)$$

where we can define:

$$\delta\boldsymbol{\theta} = \begin{bmatrix} \theta_c^1 - \theta_0^1 \\ \theta_c^2 - \theta_0^2 \\ \vdots \\ \theta_c^1 - \theta_0^{N-1} \\ \theta_c^1 - \theta_0^N \end{bmatrix}, \mathbf{e}^n = \begin{bmatrix} \frac{y^n - y_0}{(r_0^n)^2} \\ \frac{x^n - x_0}{(r_0^n)^2} \\ 0 \end{bmatrix}, \mathbf{G}_\theta = \begin{bmatrix} -(\mathbf{e}^1)^T \\ -(\mathbf{e}^2)^T \\ -(\mathbf{e}^3)^T \\ \vdots \\ -(\mathbf{e}^N)^T \end{bmatrix} \quad (9.9)$$

Giving the final representation in matrix form as.

$$\delta\boldsymbol{\theta} = \mathbf{G}_\theta \delta\mathbf{x} + \tilde{\boldsymbol{\epsilon}}_\theta^n \quad (9.10)$$

For an over-determined, full-rank system where $N > 2$ a least-squares solution can be computed for the corrections to the nominal position using,

$$\delta\hat{\mathbf{x}} = (\mathbf{G}_\theta^T \mathbf{G}_\theta)^{-1} \mathbf{G}_\theta^T \delta\boldsymbol{\theta} \quad (9.11)$$

where,

$$\delta\hat{\mathbf{x}} = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}$$

9.3.1.3 Combined Lateralation and Angulation

By combining the lateralation least-squares equations from Chapter 3 with the angulation method above, a formulation can be made for a combined ranging (lateralation) and AOA measurements (angulation) representation of the least-squares position estimation.

$$\mathbf{G}_c = \begin{bmatrix} (-\mathbf{I}^1)^T & 1 \\ (-\mathbf{I}^2)^T & 1 \\ (-\mathbf{I}^3)^T & 1 \\ \vdots & \vdots \\ (-\mathbf{I}^K)^T & 1 \\ (-\mathbf{I}^1)^T & 0 \\ (-\mathbf{I}^2)^T & 0 \\ (-\mathbf{I}^3)^T & 0 \\ \vdots & \vdots \\ (-\mathbf{I}^N)^T & 0 \end{bmatrix} \quad \begin{bmatrix} \delta\rho \\ \delta\boldsymbol{\theta} \end{bmatrix} = \mathbf{G}_c \begin{bmatrix} \delta\hat{\mathbf{x}} \\ \delta\hat{b} \end{bmatrix}$$

$$\begin{bmatrix} \delta \hat{\mathbf{x}} \\ \delta \hat{b} \end{bmatrix} = (\mathbf{G}_c^T \mathbf{G}_c)^{-1} \mathbf{G}_c^T \begin{bmatrix} \delta \boldsymbol{\rho} \\ \delta \boldsymbol{\theta} \end{bmatrix} \quad (9.12)$$

The geometry matrix above consists of satellite LOS unit vectors for satellites 1 through K (3.8b), and LOS unit vectors for angulation measurements 1 through N . The angular measurement differences $\delta\theta$ are then added to the solution alongside the pseudorange differences $\delta\rho$. Following, this representation allows for the iteration on a final state by adding successive $(\delta\hat{\mathbf{x}}, \delta\hat{b})$ results to the current receiver location estimate $(x, y, z, \delta t_u)$, as in a normal over-determined least-squares iteration estimation.

9.3.2 Deterministic Estimation Using Nearest Neighbor

When there is an abundance of measurements, it may be important for computational or power savings if the RF channels incorporate measurements on an incremental basis, until a desired estimated accuracy boundary is reached. In distributed sensor network systems, it is desirable to balance the information contribution of individual sensors (RF receivers in the current discussion) against the cost of communicating with them. In addressing this issue, various machine-learning algorithms have been applied [35, 36], where one of the simplest is known as the nearest neighbor (NN) method (also referred to as k -NN, where k is the number of measurements included). This method requires a range estimate such as supplied by one of the methods described in this chapter. The ranging likelihood function is donut-shaped for a range estimate or ridge-shaped for differential range (TDOA) as shown in Figure 9.13. In a k -NN location estimation of a target, a leader sensor node adds measurements to its own based on the closest sensor node whose measurements have not yet been incorporated. Measurements are incorporated by taking the product of the likelihood functions of the sensor nodes. An example is presented in [37], where several sensors are placed along a line with one sensor placed in a corner. Sensor locations are marked by “o” symbols and sensors used in the location estimate have a “*” overlaid on the sensor locations. Figure 9.14(a–e)

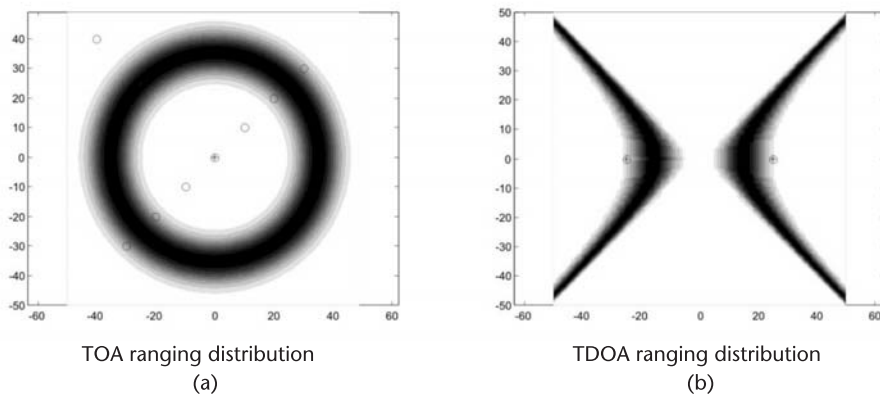


Figure 9.13 (a, b) Gaussian and differential range uncertainty.

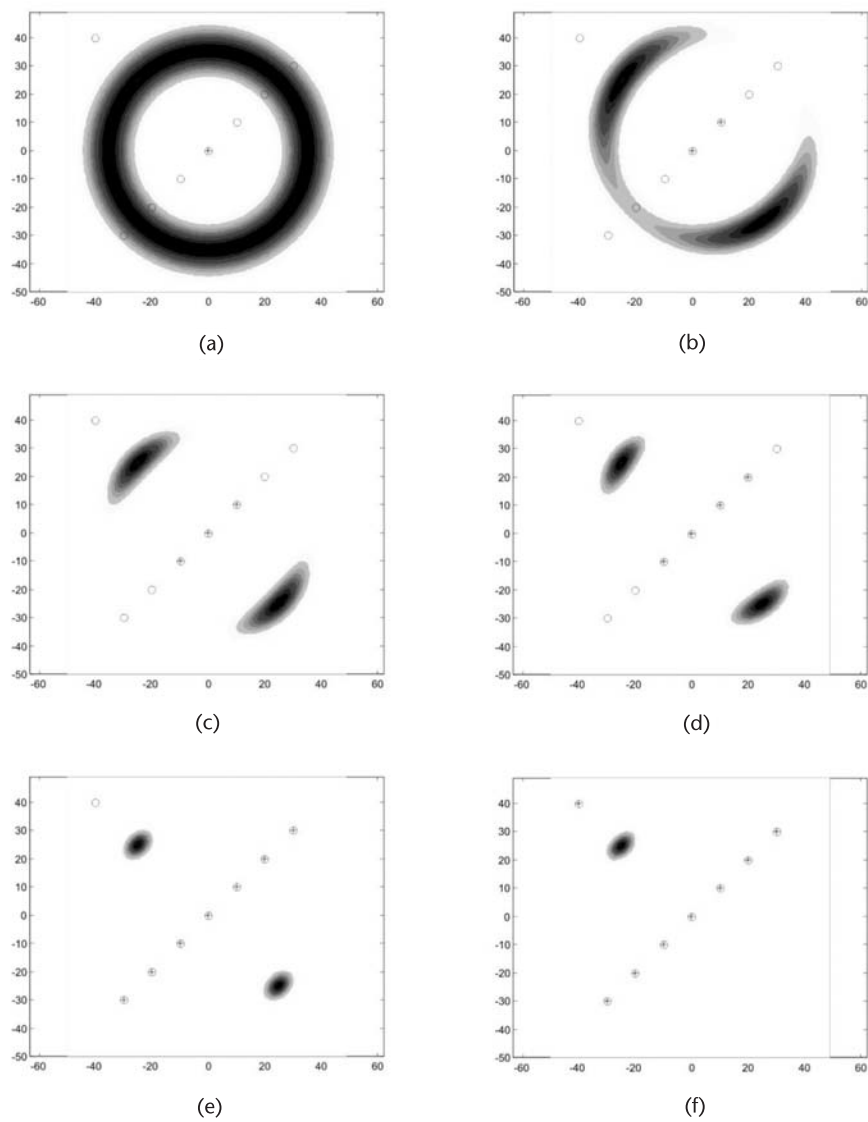


Figure 9.14 Sensor selection effect on estimate uncertainties: (a) is only the center sensor and (b)–(e) show subsequent additions of sensors along the linear axis. (f) shows the impact of the addition of the corner sensor on the ability to discern “mirror” images.

shows that as each additional sensor node is added, the distribution becomes bimodal and remains so until the corner sensor [Figure 9.14(f)] is added, thus revealing the location of the target. This example is designed to demonstrate one of the critical limitations of this approach.

A variation on the k-NN is information-driven sensor querying where some utility measure is used to choose the next sensor node. One utility function is the Mahalanobis distance, which is defined from \mathbf{y} to $\boldsymbol{\mu}$ normalized by the uncertainty covariance $\boldsymbol{\Sigma}$

$$(\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \quad (9.13)$$

and the utility function versus the current target location estimate, $\hat{\mathbf{x}}$, is defined by the negative Mahalanobis distance from the j th sensor location, $\boldsymbol{\zeta}_j$:

$$\hat{\phi}(\boldsymbol{\zeta}_j, \hat{\mathbf{x}}, \boldsymbol{\Sigma})^T = -(\boldsymbol{\zeta}_j - \hat{\mathbf{x}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\zeta}_j - \hat{\mathbf{x}}) \quad (9.14)$$

Applying this to the prior example chooses the corner sensor node far sooner than the Euclidean distance measurement. While the Mahalanobis distance and other measures like k-mean clusters [38] offer good node selection when sensors are roughly equidistant from the target, it is clear from a simple inspection of (9.14) that sensors at highly disparate distances (such as building-scale WiFi access points and GPS satellites in Earth orbit) would not provide meaningful results. In GNSS literature the most well-known geometrical measure is DOP, which is based on the covariance of the location estimate and is described in detail in Chapter 3.

Two algorithms are proposed in [39] to select the best four sensors for a position solution. The details of these methods are beyond the scope of discussion in this chapter. However, in summary, one algorithm selects three sensors based on the longest-range components in the horizontal plane, and the fourth is selected to minimize the GDOP. The second algorithm is an optimal solution that explicitly calculates the minimum GDOP based on combinations of all sensors.

9.3.3 Nonranging-Based Location Estimation

The previous examples highlight techniques useful when TOA and TDOA ranging are available. However, many systems proposed today use other approaches that result in measurements or inferences that are not based in the timing domain but rather signal power levels or some other metric. There are numerous techniques for dealing with these types of estimates.

One set of techniques that have been applied to ranging as well as signal-strength pattern matching is neural networks. In an environment that has complex geometry, creating a very complex signal pattern and power is not a simple function of distance. Modeling of the environment can be extremely involved if not impossible (or time-varying), and establishing a signal map from sampling the space can be equally difficult and subject to changes over time.

Neural network models are composed of simple interconnected elements (neurons) operating in parallel. The overall function of the network is largely determined by the connection neurons. A neural network can be trained to perform a complex pattern-recognition function by adjusting the connection between neurons [40]. The nonlinear transformation for each neuron and a sufficiently large number of free parameters guarantees that a neural network is capable of representing the relationship between inputs (signal strengths and ranges) and outputs (position). A user may train and use the method without asking for the detailed location of the access point, an important factor when training a network where this information is not known a priori [23, 41]. Results in [23] suggest that substantial improvements,

exceeding 25%, over linear-estimation and nearest-neighbor methods would be possible with the application of neural network algorithms.

Another approach called the hill-climbing search has been applied to signal-strength-only measurement applications described in [42]. The purpose of the algorithm is to minimize the signal strength error relative to empirical data. Numerous other methods such as support vector mapping (SVM), hamming distance, K-mean clustering, ratio-nearest neighbor, and others have also been put forward in the research literature. These, approaches are summarized in [26].

9.3.4 Probabilistic Estimation Using Centroid/Center of Mass

Computing based on access points, cell-tower, or any other RF radio beacon is both fast to compute and in its basic form does not employ a radio propagation model [10]. This approach uses a lookup table of cell ID locations, which include latitude and longitude entries. The centroid algorithm then estimates the receiver's position to be the geometric center of all the RF transmitters that are seen in a set of measurements. Weighting by the received signal strength observed in the measurement data is a variation that can offer a small improvement in accuracy based on a "center of mass." This approach requires precise knowledge of the location of the transmitters or at least a very good estimate based on surveying the transmitter broadcast region. Results in [30] used the centroid algorithm in its basic form without any modeling of radio propagation. It was shown in [30] that the centroid method using GSM cell-tower signals and no radio propagation modeling produced a 50% positioning error of 232m and 90% positioning error of 574m in a downtown area with a cell-tower density of 66 cells/km². In a suburban area with a cell-tower density of 26 cell/km² had positioning errors of 760 and 2,479m for 50% and 90% levels respectively.

9.3.5 Bayesian State Estimation

In a Bayesian formulation, the objective is to obtain a good estimate of state from the measurement history. This approach forms the basis of many types of filtering methods. If $p(\mathbf{x})$ represents the a priori probability distribution function (PDF) about the state \mathbf{x} , $p(\mathbf{z} | \mathbf{x})$ is the likelihood function of \mathbf{z} given \mathbf{x} (also called the sensor model), and $p(\mathbf{x} | \mathbf{z})$ is the a posteriori distribution of \mathbf{x} given the measurement \mathbf{z} (also called the state probability density). Ultimately the idea is to build a sensor model to predict the signal strength at each location and then use this information to compute the likelihood of measurements. The state probability density, $p(\mathbf{x} | \mathbf{z})$ is given by Bayes' theorem:

$$p(\mathbf{x} | \mathbf{z}) = \frac{p(\mathbf{z} | \mathbf{x})p(\mathbf{x})}{\int_{R_{\mathbf{x}}} p(\mathbf{z} | \mathbf{x})p(\mathbf{x}) d\mathbf{x}} = \frac{p(\mathbf{z} | \mathbf{x})p(\mathbf{x})}{p(\mathbf{z})} \quad (9.15)$$

where $p(\mathbf{z})$ is known as the marginal PDF and is integrated over the entire range of \mathbf{x} , $R_{\mathbf{x}}$. Since the distribution of $p(\mathbf{x} | \mathbf{z})$ is determined by the numerator, the

denominator can be treated as a normalizing function after the numerator is computed [37].

$$p(\mathbf{x} | \mathbf{z}) = \kappa \cdot p(\mathbf{z} | \mathbf{x})p(\mathbf{x}) \quad (9.16)$$

Under normal circumstances, measurements \mathbf{z} will be acquired over time from one or more receivers and are denoted by $\mathbf{z}_t = z_1, z_2, \dots, z_t$ for the current location state \mathbf{x}_t at time t . Bayes' estimation assumes that the process is Markov where sensor measurements depend only on a receiver's current physical location and that a receiver's location at time t depends only on the previous state \mathbf{x}_{t-1} . States before \mathbf{x}_{t-1} provide no additional information. So the new state probability density $p(\mathbf{x}_t | \mathbf{z}_t)$ can be updated by Bayes' rule in two steps.

Step 1:

$$p(\mathbf{x}_t | \mathbf{z}_{t-1}) = \int_{R_{\mathbf{x}_{t-1}}} p(\mathbf{x}_t | \mathbf{x}_{t-1})p(\mathbf{x}_{t-1} | \mathbf{z}_{t-1}) d\mathbf{x}_{t-1} \quad (9.17)$$

Step 2:

$$p(\mathbf{x}_t | \mathbf{z}_t) = \kappa_t \cdot p(\mathbf{z}_t | \mathbf{x}_t)p(\mathbf{x}_t | \mathbf{z}_{t-1}) \quad (9.18)$$

The function, $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ describes the system dynamics, in other words, how the location state changes over time. κ_t is a normalizing constant. In step 1 the filter predicts the current state based on the system dynamics and previous measurements. In step 2 the filter corrects the estimate based on an updated measurement. A Bayes filter of this form is an abstract approach in that it only provides for a probabilistic framework for recursive state estimation. There are multiple different implementations of Bayes filters that differ in the sensor model used, system dynamics, and state probability densities. The Kalman filter is an example of a Bayes estimator where the state probability density and sensor error models are Gaussian distributions and the system dynamics model is linear. A more detailed explanation of the Kalman filter can be found in Chapter 6 or [43]. However, many real world problems are unsuited to such assumptions. Other approaches that are variants of the Bayes estimator are: (1) multiple hypothesis tracking (MHT); (2) topological; (3) grid-based; and (4) particle filter approaches [44].

MHT uses a weighted mixture of Gaussian distributions for the state probability density function. Each hypothesis is separately expressed with a Kalman filter and the weights are determined from how well the hypothesis' estimate predict the receiver measurements. Since this approach uses multiple Kalman filters running in parallel it is computationally more demanding than a traditional Kalman filter but it is less susceptible to the modeling errors.

Topological approaches use a symbolic representation of the states to simplify computational complexity—for example, if an indoor environment location can be represented by room number and specific hallways. The dynamics model is set to how a receiver moves through the space and the sensor model needs only isolate

to the desired resolution level. The advantage of this approach is computational efficiency at the expense of coarseness of estimation.

In a grid-based approach the state probability and sensor models are mapped to a discrete grid. The predicted state probability is computed by the convolution of the original state probability with the dynamics model. The limitation of this approach is the trade-off between computational and storage capacity versus resolution (i.e., the precision cannot exceed the grid density).

Like the grid and MHT approaches, the particle filter (also known as the sequential Monte Carlo method) utilizes multiple samples of the state probability function. In contrast to the grid method, the samples are not chosen from a physical grid but are random samples or particles. The particles are weighted by a normalized importance function. One implementation of a particle filter is given in [37]. More information on the particle filter is included in Appendix 9B on the DVD.

9.4 Integration Methods

Combinations of location methods such as GNSS and signal strength as well as GNSSs and proximity do not offer direct integration schemes. It also may not be practical or possible in other cases to directly combine GNSS and other RF systems because the measurements are not available at the “black box” output. This section presents a discussion for integration at a position/velocity level. Figure 9.15 presents a general diagram illustrating the process.

9.4.1 Least-Squares Integration

The simplest combination of general navigation systems is a weighted least-squares of m independent solutions [45]. This approach is relatively simple and places low computational load on the system processor. It is noted in [45] that this approach is poorly suited for INS, IMU, and dead-reckoning integration schemes because it does not allow for the compensation of systematic biases between the navigation systems, and in the dead reckoning case, the increase in uncertainty with time just causes the sensor to be deweighted. This is not as important when considering

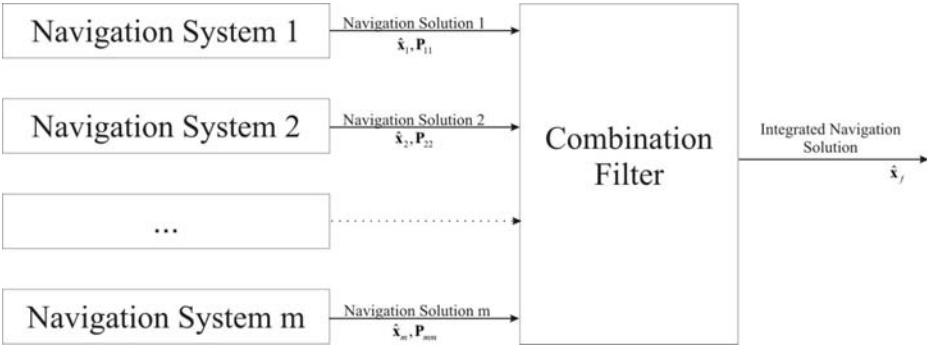


Figure 9.15 Navigation solution combination.

computing a combination of multiple RF systems. Another potential drawback is that there is no capability for integrating at different time intervals, so if the navigation systems are at very different update rates there is no state update to a common time frame. This suggests a Kalman filter approach as an alternative.

9.4.2 Kalman Filter Integration

The integration of multiple navigation solutions using a Kalman filter can also be accomplished using the state estimates from individual solutions as the measurement vector. A detailed explanation of the steps involved can be found in [45]. The combination filter is not limited to a Kalman filter implementation of a Bayesian estimator. Different approaches could be developed that would better capture some nonlinear variation in the individual location estimates. One such alternative could use a particle filter as described in Appendix 9B on the DVD as an estimator of the combined state.

9.4.3 Contextual Processing

Location inputs may be represented symbolically or descriptively in nature in contrast to fixed grid points. Different technologies employed will have dramatically different accuracies and applicable ranges leading to unbalanced weighting in least-squares integration or potentially nonlinear effects for a Kalman filter. An approach is outlined in [46] along with a corresponding processing capability for integrating a broad variety of sensors including the RF approaches of RFID tags, active RF proximity devices, GPS, and 802.11b. These disparate sensors are combined in their “fusion layer” using Bayesian filtering techniques, specifically particle filters. This toolbox uses generalized particles filters in seven dimensions (x, y, z, roll, pitch, yaw, and linear velocity) with the flexibility of reducing the dimensions to two dimensions for higher computational performance.

Other approaches to contextual processing are discussed in [26], which blends symbolic/descriptive locations (i.e., “Elm Street” or “Room 203”) with value-based spatial information. Contextual processing in those cases requires mapping multiple inputs to a common framework before combinations can be performed.

9.5 Example Systems

Previous sections have described the basic tenets of location determination followed by various methods for combining non-GNSS, RF-based location technologies with GNSS. This section will show specific examples of systems that have combined other approaches with GNSS.

9.5.1 Pseudolites

Probably the earliest combination of non-satellite RF signals with GNSS (specifically GPS) was the use of pseudo-satellites, or pseudolites, employed during the original GPS constellation testing and verification stages. In September 1978 an F4 aircraft

flew a simulated precision approach using a combination of a single GPS satellite and three ground-based pseudolites. As shown in Figure 9.16, the ground-based transmitters formed an *inverted range* in which a combination of pseudolites and navigation satellites could be employed. The pilots were provided a display that showed vertical and horizontal deviation from a desired path. In these four flights they showed that this satellite/ground architecture could provide accuracy better than 50m for vertical guidance and substantially better than 100m for horizontal guidance [47]. The solution for such a system treats the signals on the ground as GPS satellites and solves the position and clock offsets using the least-squares estimation method described in Chapter 3.

The main disadvantages of the approach is that these pseudolite transmitters broadcast signals that may block the existing satellite transmissions, especially when the receiver is much closer to the pseudolite and the relative signal strength is overwhelming the input dynamic range of the receiver (the near-far problem). The first to point out the advantages of pseudolites beyond that of test platforms was [48], where a method of pulsing the signal was proposed to address the near-far problem. Further progress was introduced in ranging pseudolites over the following 20 years and selected papers are listed in the further reading section on the DVD.

Applications of pseudolites to POA systems was introduced in [49–51], and [32] describes a particular example where the large relative motion of the receiver versus a fixed pseudolite allows for population of a geometry matrix to estimate

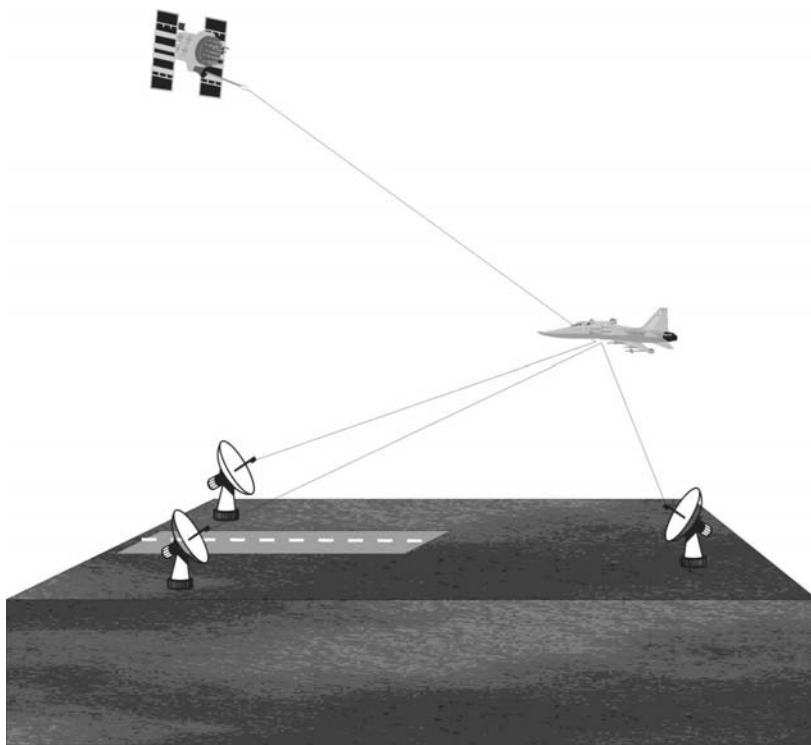


Figure 9.16 GPS Inverted range using pseudolite transmitters.

the carrier phase integers. If the LOS is captured over time, large changes to the LOS will make the integer uncertainty observable. In these cases, geometry plays a critical role in eliminating false integer combinations and enabling centimeter-level location accuracy.

9.5.2 Synchrolites

Also described in [32] is the concept of the *synchronized pseudolite* or *synchrolite*, which synchronizes its signal using one or more GPS satellite signals. This concept was employed for indoor application in [52] and solved two problems: (1) The GPS signal could be rebroadcast indoors where signals may not be available directly from the satellites; and (2) it provided multiple broadcast sites with timing delays to allow TDOA measurements for indoor positioning. The multiple broadcast sites are important because retransmission of GPS signals from a single repeater would give the location of the repeater antenna at the indoor receiver.

The delays in the broadcast antennas are selected to be at least one GPS chip (293m) so that the received signals suffer from reduced interference because of the design of the modulation codes. An overview of a system of this type is depicted in Figure 9.17, which permits the usage of standard GPS tracking hardware. An implementation of this system produced a two-dimensional RMS error less than 2m in a 20-m room [52].

9.5.3 Self-Synchronizing Networks

An extension of the pseudolite concept called the “LocataLite” has been formulated and developed by the Locata Corporation of Australia. Instead of using the GPS L1 frequency, this system uses the industrial, scientific, and medical (ISM) 2.4-GHz band for its channels. The basis of the system is the LocataLite, a self-synchronizing

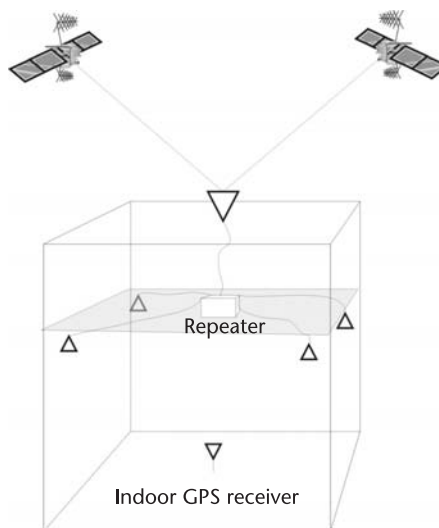


Figure 9.17 Indoor GNSS repeater for TDOA indoor locations.

transceiver that transmits with a chipping rate of 10 MHz. Multiple LocataLites are synchronized with one another down to a level of 10 picoseconds to form a LocataNet, which allows all devices to share a common time base. By using a minimum of four LocataLites an autonomous location system can be formed to locate a receiver using carrier phase [53]. In order to reference the local-area coordinate frame formed by the LocataNet to WGS-84 (the GPS coordinate frame) a self-survey can be formed at one or more of the LocataLites to coordinate the timing and position of the LocataNet to GPS. Accuracy of this system has been reported as good as 2 cm RMS outdoors and 20 cm RMS indoors [54].

9.5.4 GPS and Relative Navigation

A variation of TOA/TDOA location is when locations are determined not against some fixed reference (such as the WGS-84 in GPS) but between various transceivers in a system. This type of system uses a relative position between two or more nodes that are in communication with at least one node at a known location. One operational example of such a system is the joint tactical information distribution system (JTIDS). JTIDS is a synchronous, jam-resistant, spread-spectrum, secure communication, navigation and identification system. In these cases each transmitter in the network will occasionally transmit its own position and velocity estimate (along with the uncertainties of position and time estimates) with signals in the 960–1,215-MHz band. A receiver in the network receives these messages and passively ranges to the transmitter using the TOA and position messages. As described in [55], a receiver based on these ranging measurements would normally be paired with an inertial measurement dead-reckoning unit in order to provide continuous location estimates between the 3- to 12-second ranging updates supported by the system. Like GPS, GDOP applies to JTIDS as an indicator of the relative geometry quality between the receiver and transmitters, both operating in L-band. JTIDS has the advantage that since the operational range for the system is only a few hundred kilometers (significantly less than the transmit ranges of GNSS), the signal power requirements are eased. This system sees its greatest advantage when the number of signals for GPS is insufficient for stand-alone operation, where JTIDS offers a reliable ground-based alternative. Conversely, when LOS JTIDS signals are blocked or unavailable due to limited number of broadcasting transceivers, GPS signals may be used as a backup.

9.5.5 TV-Based Location

One of the most challenging aspects of GNSS navigation occurs when trying to navigate inside of buildings and urban environments, using weak and reflected signals (See Chapter 12). Many systems designed to augment and complement GNSS seek to specifically address this issue. Television signals are broadcast all over the world. These channels contain significantly more power, over 40 dB more compared to GNSS signals at the surface of the Earth. Additionally, with television signals being terrestrially based, the source of LOS blockage would likely be uncorrelated to blockage due to satellite signals or at least present a different type of blockage compared to GNSS signals (e.g., pass-through wall versus pass-through

ceiling). A major advantage of television signals to alternative signals is that the number of receivers for a given transmitter is unlimited. Television broadcasts are at a lower frequency than GNSS signals giving better structure penetration and transmit at a relatively high bandwidth of 6 MHz providing acceptable phase resolution (see Appendix 9A on the DVD) [56].

As illustrated by Figure 9.18, a local monitor station (with a reference receiver) measures the signal timing of the television signals and GNSS signals and reports this to a location server. The receiver to be located measures the TOA of the television and GPS signals and forwards this to the location server to compute a range-based solution. In a system implementation, combined TV+GPS location accuracies in a highly challenging urban environment have been reported in below 50-m CEP [57].

9.5.6 Integration of Cellular Location Systems and GNSS

The FCC “E-911” mandate in 1996 greatly increased the requirement on cell phone location knowledge in the United States during an emergency call and spurred a significant amount of academic and commercial research into a variety of cell-phone location technologies as well as augmentations to GPS sensitivity. One method for cell phone location was developed for phones using CDMA technology called advanced forward link trilateration (AFLT). This technique is based on the *forward link*, or the path from the cell tower to the phone, and uses measurements at the handset receiver to form pseudoranges between the tower and phone. This succeeded because CDMA signals are synchronized to a common time base (currently GPS). Measurements are then collected at the handset and sent to a location server through the cell phone messaging system where the pseudoranges are

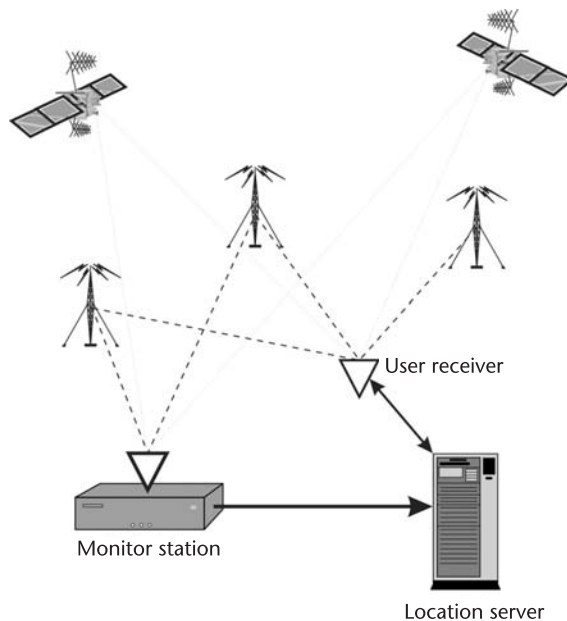


Figure 9.18 Overview of a television positioning system.

combined into a location estimate. A minimum of three signals from cell towers are necessary to estimate the two-dimensional location of the handset along with the clock uncertainty. The advantage of using cellular signals is that compared to GPS, cell signals are received at power levels that are 60-dB higher. This allows for substantially increased availability of location solutions in a variety of challenging environments such as indoors where cell signals may be received but GNSS signals are unavailable.

As CDMA signals are synchronized to GPS, it was natural for hybrid approaches combining GPS and AFLT to surface. Under normal operations GPS signals and AFLT are received at the handset receiver. All measurements are sent to a location server, which has the location of the cell tower or towers and the GPS satellite ephemeris, and a location is estimated. Results for a combination of AFLT and GPS on a live network showed two-dimensional RMS errors of 6.28m in a nonchallenging environment [58]. The authors indicated that the lower performance is generally likely due to multipath or LOS blockages.

In contrast to CDMA cell phone networks, the GSM cell technology does not synchronize its signals to GPS or GNSS. The frequency stability of GSM cell towers is fairly high however, exceeding 0.05 ppm, and this allows for clock errors to be measured and modeled as in the case for analog TV signals. In [59] a system was developed for estimating GSM cell tower clock offset using GPS and then taking the resulting clock biases to determine pseudorange information between a handset and a cell tower. Results were presented for two-dimensional locations of a handset using both multiple GPS satellites plus multiple GSM TDOA. The low resolution of timing measurements on the GSM signal limited the performance to more than 800m in two-dimensional RMS with either one or two GPS satellites. This improved to under 600m two-dimensional RMS when a TOA measurement was combined with two GPS pseudoranges. While the results of [59] show that some advantage of timing measurements could be achieved on a GSM system, results in [28, 30] suggest that approaches for GSM networks based on signal-strength measurements may be substantially better.

9.6 Examples Included on the DVD

This section includes example calculations from the text as well as some helpful utilities for RF-based location determination in general.

9.6.1 RF Antennas

The DVD presents an Excel spreadsheet that allows for basic far-field antenna efficiency computations over a range of RF frequencies.

9.6.2 Doppler Calculations

The DVD provides Matlab/Octave code that gives a simulated example of using Doppler measurements for instantaneous and multiepoch location determination.

9.6.3 K-Nearest Neighbor Plot

The DVD includes Matlab/Octave code that computes the example k-nearest neighbor estimation plots for different sensors.

9.7 Further Reading

More details on some of the subjects discussed in this chapter, as well as a comprehensive list of references, have been included on the DVD for those readers who wish to explore the subject of combining GNSS and RF systems in more detail.

References

- [1] Mannings, R., *Ubiquitous Positioning*, Norwood, MA: Artech House, 2008.
- [2] Hightower, J., and G. Borriello, "Location Systems for Ubiquitous Computing," *IEEE Computer*, Vol. 34, Issue 8, IEEE Computer Society Press, 2001, pp. 57–66.
- [3] Kraus, J. D., and R. J. Marhefka, *Antennas: For All Applications*, New Delhi India: Tata McGraw-Hill Publishing Company Limited, 2003.
- [4] Enge, P., et al, "Terrestrial Radionavigation Technologies," *Navigation: Journal of The Institute of Navigation*, Vol. 42, No. 1 Special Issue, Institute of Navigation, 1995.
- [5] Jenkins, H. H., *Small-Aperture Radio Direction-Finding*, Norwood, MA: Artech House, 1991.
- [6] Kolodziej, K., and J. Hjelm, *Local Positioning Systems: LBS Applications and Services*, Boca Raton, FL: CRC Press Taylor & Francis Group, 2006.
- [7] Gollu, A., "RTLS and Yard Management: Increasing Visibility," PINC Solutions presentation at the Wireless Communications Alliance RFID SIG, http://www.wca.org/event_archives/2007/GOLLU_RFID_SEPT2007.pdf, September 2007.
- [8] Bargh, M. S., and R. de Groote, "Indoor Localization Based on Response Rate of Bluetooth Inquiries," *The First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments: MELT-2008*, September 19, 2008, <http://www2.parc.com/isl/projects/MELT08/program.htm>.
- [9] Aksu, A., J. Kabara, and M. B. Spring, "Reduction of Location Estimation Error Using Neural Networks," *The First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments: MELT-2008*, <http://www2.parc.com/isl/projects/MELT08/program.htm>, September 19, 2008.
- [10] LaMarca, A., et al, "Place Lab: Device Positioning Using Radio Beacons in the Wild," *Proceedings of Pervasive 2005*, Munich, Germany.
- [11] Kayton, M., and W. R. Fried, *Avionics Navigation Systems*, New York, NY: John Wiley & Sons, Inc., 1997.
- [12] Fuller, R. A., *Aviation Utilization of Geostationary Satellites for the Augmentation to GPS: Ranging and Data Link*, Ph.D. thesis, Department of Aeronautics and Astronautics, Stanford University, CA, May 2000.
- [13] Helfrick, A., "Apparatus and Method for Coupling Satellite Navigation Signals to a VOR System," U.S. Patent 5,801,659, September 1998.
- [14] Enge, P., and P. Misra, *Global Positioning System, Signals Measurements and Performance*, Lincoln, MA: Ganga-Jamuna Press 2004.
- [15] Guier, W. H., and G. C. Weiffenbach, "Genesis of Satellite Navigation," *Johns Hopkins Appl. Technical Digest*, Volume 19, Number 1, 1998.

- [16] Parkinson, B., et al, "A History of Satellite Navigation," *Navigation: Journal of The Institute of Navigation*, Vol. 42, No. 1 Special Issue, Institute of Navigation, 1995.
- [17] Kusý B., À. Lêdeczi, and X. Koutsoukos, "Tracking Mobile Nodes Using RF Doppler Shifts," *ACM 5th Conference on Embedded Networked Sensor Systems (SenSys)*, Sydney, Australia, November 2007.
- [18] Lee, W. C. Y., *Wireless and Cellular Telecommunications* (Third Edition), New York, NY: McGraw-Hill Companies, 2006.
- [19] Pahlavan, K., and A. H. Levesque, *Wireless Information Networks*, New York, NY: John Wiley & Sons, 2005.
- [20] Lee, W. C. Y., *Mobile Communications Design Fundamentals*, (Second Edition), New York, NY: John Wiley & Sons, 1993.
- [21] Bahl, P., and V. N. Padmanabhan, "RADAR: An In-Building RF-Based User Location and Tracking System," *Proceedings of INFOCOM 2000: Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, IEEE, 2000.
- [22] Seidel, Y. S., and T. S. Rapport, "914 MHz Path Loss Prediction Model for Indoor Wireless Communications in Multi-Floored Buildings," *IEEE Trans. on Antennas & Propagation*, Feb. 1992.
- [23] Hatami, A., *Application of Channel Modeling for Indoor Localization Using TOA and RSS*, Ph.D. thesis, Department of Electrical and Computer Engineering, Worcester Polytechnic University, Worcester, MA, May 2006.
- [24] Bargh, M., S., and R. de Groote, "Indoor Localization Based on Response Rate of Bluetooth Inquiries," *The First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments: MELT-2008*, <http://www2.parc.com/isl/projects/MELT08/program.htm>, September 19, 2008.
- [25] King, T., T. Haenselmann, and W. Effelsberg, "Deployment, Calibration, and Measurement Factors for Position Errors in 802.11-Based Indoor Positioning Systems," *Proceedings of Location-and Context-Awareness, Third International Symposium, LoCA 2007*, Oberpfaffenhofen, Germany, 2007.
- [26] Kjaergaard, M. B., "A Taxonomy for Radio Location Fingerprinting," *Proceedings of Location-and Context-Awareness, Third International Symposium, LoCA 2007*, Oberpfaffenhofen Germany, 2007.
- [27] Ji, Y., et al, "ARIADNE: A Dynamic Indoor Signal Map Construction and Localization System," *MobiSys 2006: Proceedings of the 4th International Conference on Mobile Systems, Applications and Service*, ACM Press, 2006.
- [28] Laitinen, H., J. Lahtenmaki, and T. Nordstrom, "Database Correlation Method for GSM Location," *Vehicular Technology Conference, VTC 2001*, IEEE VTS 53rd, Vol. 4 (2001), pp. 2504-2508, 2001.
- [29] Madigan, D., et al, "Bayesian indoor positioning systems," *Proceedings IEEE INFOCOM 2005, 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, IEEE, 2005.
- [30] Chen, M., et al, "Practical Metropolitan-Scale Positioning for GSM Phones," *Proceedings of Ubicomp 2006*.
- [31] Hofmann-Wellenhof, B., H. Lichtenegger, and J. Collins, *Global Positioning Systems Theory and Practice*, Fifth Revised Edition, New York: Springer-Verlag, 2001.
- [32] Cobb, H. S., *GPS Pseudolites: Theory, Design, and Applications*, Ph.D. thesis, Department of Aeronautics and Astronautics, Stanford University, CA, 1997.
- [33] Federal Communications Commission (FCC), "First Report and Order in The Matter of Revision of Part 15 of the Commission's Rules Regarding Ultrawideband Transmission Systems," ET-Docket 98-153, FCC 02-048, released April 2002.
- [34] Gavish, M., and A. J. Weiss, "Performance Analysis of Bearing-Only Target Location Algorithms," *Aerospace and Electronic Systems, IEEE Transactions*, Volume: 28 Issue: 3, IEEE, 1992.

- [35] Witten, I. H., and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edition, San Francisco: Morgan Kaufmann, 2005.
- [36] Crassidis, J. L., and J. L. Junkins, *Optimal Estimation of Dynamical Systems*, Boca Raton, FL: Chapman & Hall/CRC Press, 2004.
- [37] Zhao, F., and L. Guibas, *Wireless Sensor Networks: An Information Processing Approach*, San Francisco, CA: Morgan Kaufmann Publishers, 2004.
- [38] Otsason, V., et al, "Accurate GSM Indoor Localization," *Proceedings of Ubicomp 2005*, Tokyo, Japan. September 2005.
- [39] Noe, P. S., K. A. Myers, and T. K. Wu, "A Navigation Receiver Algorithm for the Low-Cost GPS Receiver," *Global Positioning System, Volume 1*, Institute of Navigation, Washington D.C., 1980.
- [40] Hagan, M., H. Demuth, and M. Beale, *Neural Network Design*, Oklahoma City OK: PWS Publishing Company, 1996.
- [41] Battiti, R., A. Villani, and T. L. Nhat, "Neural Network Models for Intelligent Networks: Deriving the Location from Signal Patterns," *Proceedings of the First Annual Symposium on Autonomous Intelligent Networks and Systems*, UCLA, Los Angeles, CA, 2002.
- [42] Hightower, J., G. Borriello, and R. Want, "SpotON: Indoor location sensing based on RF signal strength," UW SCE Technical Report #2000-02-02, February 2000.
- [43] Simon, D., *Optimal State Estimation: Kalman, H Infinity and Non-Linear Approaches*, New York, NY: John Wiley & Sons, 2006.
- [44] Fox, D., et al, "Bayesian Filtering for Location Estimation," *IEEE Pervasive Computing*, Vol. 2, No. 3, IEEE Computer Society Press, July–September 2003.
- [45] Groves, P. D., *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*, Norwood, MA: Artech House, 2008.
- [46] Hightower, J., G. Borriello, and D. Fox, "The Location Stack," *Intel Research Seattle Report IRS-TR-03-008*, July 2003.
- [47] Strada, J. A., and D. W. Henderson, "Navstar Field Test Results," *Global Positioning System, Volume 1*, Institute of Navigation, Washington D.C., 1980.
- [48] Klein, D., and B. W. Parkinson, "The Use of Pseudolites for Improving GPS Performance," *Global Positioning System, Volume 3*, Institute of Navigation, Washington, D.C., 1986.
- [49] Pervan, B. S., C. E. Cohen, and B. W. Parkinson, "Autonomous Integrity Monitoring for Precision Approach using DGPS and a Ground-Based Pseudolite," *Proceedings of ION GPS '93*, Institute of Navigation, Salt Lake City, Utah, September 1993.
- [50] Cohen, C. E., et al, "Real-Time Flight Test Evaluation of the GPS Marker Beacon Concept for Category III Kinematic GPS Precision Landing," *Proceedings of ION GPS '93*, Institute of Navigation, Salt Lake City, Utah, September 1993.
- [51] Cohen, C. E., et al, "Real-Time Cycle Ambiguity Resolution Using a Pseudolite for Precision Landing of Aircraft with GPS," presented at the *Second International Symposium on Differential Satellite Navigation Systems*, Amsterdam, the Netherlands, March 1993.
- [52] Im, S. H., G. I. Jee, and Y. B. Cho, "An Indoor Positioning System Using Time-Delayed GPS Repeater," *Proceedings of the ION GNSS 19th International Technical Meeting of the Satellite Division*, Institute of Navigation, Fort Worth, TX, 2006.
- [53] Barnes, J., et al, "High Accuracy Positioning Using Locata's Next Generation Technology," *Proceedings of the Institute of Navigation Conference (ION GNSS-2005)*, Institute of Navigation, Long Beach CA, 13–16 September 2005.
- [54] Barnes, J., Rizos, C., Wang, J., Small, D., Voigt, G., Gambale, N., "LocataNet: A new positioning technology for high precision indoor and outdoor positioning," *Proceedings of the 16th Int. Tech. Meeting of the Satellite Division of the U.S. Institute of Navigation*, Portland Oregon, 9–12 September 2003.
- [55] Fried, W. R., "Operational Benefits and Design Approaches for Combining JTIDS and GPS Navigation," *Global Positioning System, Volume 3*, Institute of Navigation, Washington, D.C., 1986.

- [56] Rabinowitz, M., and J. J. Spilker, "The Rosum Television Positioning Technology," *Proceedings of the ION 59th Annual Meeting/CIGTF 22nd Guidance Test Symposium*, Albuquerque, NM, 23–25 June 2003.
- [57] Do, J. Y., M. Rabinowitz, and P. Enge, "Performance of Hybrid Positioning System Combining GPS and Television Signals," *Proceedings of Position, Location, And Navigation Symposium 2006*, IEEE/ION, San Diego, CA, 2006, pp. 556–564.
- [58] Kim, H. S., et al, "Performance Analysis of Position Location Methods Based on IS-801 Standard," *Proceedings of ION GPS 2000*, Salt Lake City, UT, 19–22 September 2000.
- [59] Favey, E., D. Ammann, and C. Bürgi, "Hybrid Positioning using GPS and GSM Ranging Measurements," *Proceedings of ION GNSS 17th International Technical Meeting of the Satellite Division*, Long Beach, CA, 21–24 Sept. 2004.

Aviation Applications

Jason Rife and Sam Pullen

10.1 Introduction

Aviation users represent a small fraction of the overall market for GNSS devices, but their demanding applications continue to advance the cutting edge of satellite navigation technology. Perhaps surprisingly, the primary concern in aviation applications of GNSS is not accuracy. Though accuracy is an important factor, the most critical characteristic for design and certification of aviation GNSS receivers is reliability. These systems must essentially never introduce a spurious signal that could compromise the safety of passengers or aircraft equipment. Specialized augmentation systems are thus required to ensure safety by monitoring GNSS for spurious signals and promptly alerting pilots in the case of an anomaly.

This chapter focuses on augmentation systems as the cornerstone of aviation applications of GNSS. The first half of the chapter discusses the basic classes of augmentation systems and their applications. The second half describes quantitative techniques used to analyze augmentation system performance.

10.2 Classes of Aviation Augmentation Systems

Four classes of augmentation systems have been recognized by the international aviation community. These categories include: the aircraft-based augmentation system (ABAS), the SBAS, the GBAS, and a hybrid architecture known as the ground-based regional augmentation system (GRAS). The aircraft-based approach employs monitors built into user avionics and requires no external infrastructure (other than the GNSS satellites themselves). These monitors enable the construction of rigorous error bounds by detecting instances of hazardously misleading information (HMI), a term referring to any threatening GNSS anomaly (see Section 10.5.1). By comparison to the aircraft-based approach, the other classes of augmentation systems all employ an infrastructure of terrestrial reference receivers. These receiver networks enhance the sensitivity of HMI monitoring. Additionally, these networks enable the broadcast of differential corrections that significantly improve user accuracy. All four classes of augmentation systems are illustrated in Figure 10.1.

ABAS offers a distinct advantage in that it can be used nearly anywhere that GNSS satellites are in view. Although ABAS may incorporate non-GNSS sensors,

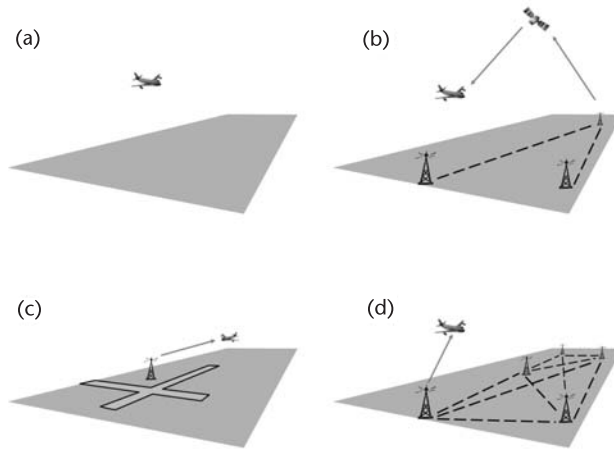


Figure 10.1 Four categories of augmentation systems: (a) ABAS, (b) SBAS, (c) GBAS, and (d) GRAS.

an important subcategory of ABAS is GNSS-only RAIM. This approach implements monitoring using the least-squares residuals from the navigation solution. A large residual corresponds to a measurement that diverges from other measurements. By excluding divergent satellite measurements from the navigation solution, RAIM detects large HMI events and thereby can establish a tighter confidence bound on the navigation sensor error. To obtain nonzero residuals, RAIM requires at least one more measurement than conventional GNSS navigation (five satellites rather than four).

SBAS monitors for HMI using a network of terrestrial receivers distributed over vast distances, with coverage areas typically on the scale of large countries or continents. Alert messages, error bounds, and differential corrections are broadcast to users in this coverage area via a space-based communications link, most typically via a satellite positioned in GEO. In the United States, the Federal Aviation Administration (FAA) introduced the world's first SBAS when it declared WAAS operational in 2003. Around the world, other governments are coordinating new SBAS implementations, such as Japan's Multifunctional-Transport SBAS (MSAS), operational in 2007; the European Geostationary Navigation Overlay Service (EGNOS), expected operational in 2009; and India's GPS and GEO Augmented Navigation System (GAGAN), expected operational in 2010. SBAS offers a distinct advantage over ABAS in that it provides differential corrections that estimate and mitigate major GNSS error sources, including ionosphere, troposphere, and clock errors. As GEOs broadcast over standard GNSS frequencies, conventional antennas can receive SBAS transmissions. In fact, each GEO provides users with an additional ranging measurement that effectively transforms the GEO into another GNSS satellite (although ranging measurements from today's GEO satellites are not as accurate as measurements from GPS satellites).

GBAS monitors for HMI using a small network of receivers distributed over short baselines, on the order of hundreds of meters. These systems, such as the FAA's LAAS [1] and the U.S. military's Joint Precision Approach and Landing System (JPALS) [2], are intended to support high-precision aviation applications

over a compact service volume (less than 60 km in radius). In GBAS, alerts and differential corrections are broadcast from a terrestrial VHF transmitter. Although the GBAS message reaches fewer users than would an SBAS message, GBAS users benefit from their proximity to the reference antenna. At short ranges, differential corrections are significantly more effective in removing spatially correlated ionosphere and troposphere errors, resulting in higher accuracy (see Chapter 4). Also, the simpler broadcast structure of GBAS (direct VHF communication rather than “bent-pipe” communication via a satellite) results in shorter communication latency. Alert times are thus shorter for GBAS than SBAS, an important factor in achieving the tight time-to-alert requirements for precision applications such as low-visibility landing.

GRAS is a hybrid that exploits widely distributed networks of terrestrial receivers, like SBAS, but communicates differential corrections to users via ground-based VHF transmitters, like GBAS. The GRAS concept has been developed in large part to enable an SBAS-like capability for Australia and the South Pacific, where political, technical, and economic factors have made GEO access difficult and where access is readily available to a pre-existing VHF transmission network [3]. GRAS is intended to support nonprecision and precision approach operations while interfacing with GBAS to support automated landing. GRAS is scheduled to reach initial operating capability in Australia in 2009 [4].

10.3 Benefits of GPS and Augmentations to Aviation Users

Prior to the advent of NAVSTAR GPS, aircraft navigation relied primarily on inertial sensors and a network of ground-based radio transmitters [5]. As soon as the FAA declared the GPS constellation operational in 1994, civil aviation was quick to adopt GNSS technologies [6, 7]. Satellite navigation signals offer a significant benefit for aviators in that they are available globally and enable a uniform quality of navigation throughout all phases of flight. Accordingly, more flexible route planning and higher-capacity operations (with tighter separation minima) are possible. GNSS signals also reduce user operating costs and permit the decommissioning of underutilized ground-based navigation aids. The benefits of GNSS and GNSS augmentations depend on the application, as discussed in Sections 10.3.1–10.3.3.

10.3.1 Oceanic Flight

Satellite navigation has been particularly beneficial in supporting transoceanic flights. Historically, the lack of terrestrial navigation aids and radar installations has significantly hindered navigation and surveillance functions for flights over water. As a consequence, aircraft spacing has depended primarily on procedures (including predefined flight paths) rather than on sensing. Separation minima have been correspondingly large, historically 60 nm in the lateral and longitudinal directions. Because satellite navigation does not require ground-based facilities, GNSS technologies are particularly well-suited for oceanic navigation. As a consequence, civil aviation authorities have certified GPS as a “primary-means” system for

navigation over the ocean (as well as for flights in other remote areas). Separation minima have already decreased from 60 nm to 50 nm for properly equipped aircraft in certain oceanic regions. Civil aviation authorities expect an eventual reduction of oceanic separation minima to 30 nm worldwide [8]. Safety guarantees are not possible, however, without signal monitoring provided by ABAS.

10.3.2 Overland Flight: En Route, Terminal, and Nonprecision Approach

Overland flights benefit from a pre-existing, ground-based communication, navigation and surveillance (CNS) infrastructure. Ground-based beacons have defined the international standard for en route navigation under instrument meteorological conditions for over half a century [9]. Examples of beacon systems include tactical air navigation (TACAN), VOR, distance measurement equipment (DME), and combined VOR and TACAN (VORTAC).

GNSS technology provides significant new capabilities to improve operational efficiency beyond what has been possible with ground-based navigation beacons, alone. For instance, GNSS navigation provides higher accuracy than ground-based beacons, on the order of tens of meters rather than hundreds. More significantly, GPS supports area navigation (RNAV), allowing for flexible flight paths that are not necessarily constrained to lie along routes between navigation aids. The FAA has certified GPS for supplemental navigation in many phases of flight (en route, terminal and nonprecision approach) and will likely certify GPS for primary means navigation in the future. Safety guarantees are not possible without ABAS or SBAS, however. ABAS is already widely used by commercial aircraft equipped with multimode receivers (MMRs), and SBAS (in the form of WAAS) is now used by private pilots as an affordable alternative to general aviation.

10.3.3 Precision Approach and Landing

Most approaches and landings occur under visual flight rules (VFR). Accurate and robust navigation technologies are nonetheless absolutely critical to enable safe landings under low-visibility conditions [instrument flight rules (IFR)]. ILS remains the predominant technology used to support aircraft landing under instrument conditions. ILS installations operate by creating a pair of signals, called the glide slope and localizer, that together allow an aircraft to determine its vertical and lateral deviations from a reference trajectory leading downward toward the runway. ILS technology has proven its reliability over decades of operation, but it is also somewhat expensive to deploy and maintain [10]. Since it allows only for straight-in approaches and not for curved approaches, ILS technology also restricts the development of new procedures that could enhance terminal area traffic flow in the future. An enhanced technology known as the Microwave Landing System (MLS) was once perceived to be the successor to ILS, but it was largely abandoned when more cost-effective GNSS-based solutions were proposed [11].

After years of research, GNSS-based landing solutions have begun to emerge. WAAS has already been certified to support some ILS-like operations, designated as localizer performance with vertical guidance (LPV) approaches. For these approaches, the pilot must descend below clouds or fog and establish visual contact

with the runway by a decision altitude of 250 ft. WAAS capabilities will be extended in the near future to enable a new type of operation called LPV-200 [12], which is similar in nature to a category-I landing operation in that it enables a decision altitude of 200 ft.

Automated landing capabilities will be provided by GBAS (by LAAS, for example). Although international deployment of LAAS would provide an enormous benefit to commercial aviation, certification of LAAS was not possible prior to 2008, in large part due to concerns about system reliability during severe ionosphere storms [13, 14]. LAAS certification for category-I landing is expected in Summer 2009. Continued research and development will be necessary to extend LAAS capabilities to handle category-II and category-III operations (100-ft and 50-ft decision altitudes, as illustrated in Figure 10.2, adapted from [15]). The desired end state for LAAS is an enhanced Category-IIIc system that will fully support automated landing and rollout under zero-visibility conditions [1].

10.4 Future of GNSS Navigation in Aviation

As the development of aviation augmentation systems like WAAS and LAAS continues, the needs of aviation users and the nature of GNSS continues to evolve. New satellite systems are being introduced worldwide. These satellites will provide a myriad of new signals that will enhance aircraft navigation but that will also motivate the development of new augmentations. In the meantime, the aviation community is exploring radically new approaches to air traffic management (ATM) that will double or triple the capacity of the worldwide airspace. An emphasis on ubiquitous and accurate positioning within this new ATM architecture will lead to an increased reliance on GNSS navigation. For this reason, and because of enhanced security concerns in the new millennium, backup navigation systems will be absolutely essential to ensure graceful degradation of navigation performance should GNSS services be interrupted.

10.4.1 GNSS Modernization

It is likely that GNSS will radically evolve between 2010 and 2020, as modernized satellites and new satellite constellations arrive in orbit. An immediate impact for

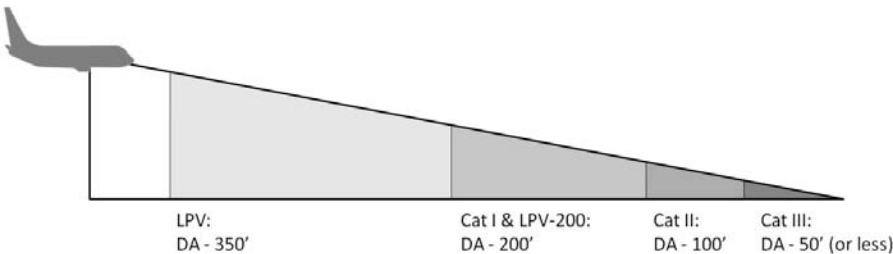


Figure 10.2 ILS-like approaches are enabled by GNSS augmentation systems.

aviation will be the improved accuracy possible with new signal structures (such as BOC modulation) and the ability to mitigate ionosphere delays using multiple frequencies [16]. In the longer term, it will be enormously beneficial to employ ranging signals from heterogeneous constellations to augment geometric diversity. Multiple constellations will provide a more even distribution of satellites across the sky, yielding more accurate and more robust navigation solutions that significantly increase the percentage of time that aviation augmentations (and especially ABAS) are available [17]. Merging ranging signals from multiple constellations is not a trivial proposition. However, given the significant potential benefits of multi-constellation navigation, it is anticipated that this hurdle will be overcome.

As new constellations are launched, GPS will remain a critical component of the overall GNSS infrastructure. A modernized GPS constellation will offer a range of new features. For aviation users, the most important of these has been the introduction, beginning with block IIF satellites, of an additional civil signal called L5C. Together the L1 and L5 civil signals can be combined for ionosphere-free navigation [18], alleviating the largest error source and the most threatening anomaly for GNSS navigation. The “magic” of the L1 and L5 frequencies is that, unlike the L2 frequency, they lie in an internationally regulated region of the electromagnetic spectrum designated for aeronautical radio navigation services (ARNS). The ARNS designation is critical to ensuring availability of signals worldwide, as radar and mobile services may interfere with the L2 frequency in some regions [4].

If all goes well, four GNSS constellations will soon offer worldwide service: GPS, operated by the United States; GLONASS, operated by Russia; Galileo, operated by Europe; and Compass/Beidou, operated by China. The Russian government is attempting a major revitalization of the GLONASS constellation following a gradual decline in its size through the late 1990s and early years of the current century. New GLONASS satellites will introduce new signals and will likely transition away from frequency domain multiple access (FDMA) in favor of code division multiple access (CDMA). Europe’s Galileo will introduce an entirely new satellite navigation constellation, with three signals lying in ARNS bands (designated E1, E5a, and E5b). In addition to an open navigation service, Galileo will also provide specialized data, including a new safety of life service. China will deploy a mixture of medium Earth orbit (MEO) and GEO satellites to evolve its current satellite navigation capability into a worldwide service called Compass/Beidou. Compass/Beidou signals are expected to be interoperable with Galileo, but few details about the proposed constellation have been released. (See Chapter 18.) In addition to these worldwide services, regional satellite systems, including Japan’s QZSS and the IRNSS, are scheduled for launch beginning in the 2010 timeframe.

As new constellations emerge, new augmentation systems will be needed. In the next decade, SBAS will be updated to provide widespread LPV-200 service and to support multiple frequencies [12]. With an eye toward the more-distant future, the FAA has commissioned the GNSS Evolutionary Architecture Study (GEAS) to consider new options for providing LPV-200 service worldwide. This panel has identified three possible options: including (1) a GNSS integrity channel (GIC) that integrates existing worldwide SBAS capabilities, (2) an approach called absolute RAIM that extends the conventional ABAS model, and (3) a hybrid approach

called relative RAIM that blends conventional SBAS and ABAS methods. Preliminary analysis indicates that these options present different cost and availability trade-offs [19]. The GIC approach appears to provide high availability using a relatively expensive infrastructure of reference stations and GEO satellites. Absolute RAIM requires minimal infrastructure but appears to deliver reduced availability. The hybrid relative RAIM approach appears to fall between these two extremes in terms of both cost and availability.

10.4.2 Next-Generation Air Traffic Management System (NextGen)

An international effort is under way to transform the practice of air traffic management to double or triple airspace capacity by 2025. In the United States, this effort is known as the “NextGen” program. GNSS navigation will be an important component of the NextGen CNS infrastructure [20]. In particular, capacity will be increased through the use of four-dimensional trajectories that designate a precisely timed path for each aircraft, from departure gate to arrival gate. If robustness issues can be resolved [21], the use of four-dimensional trajectories will enable optimized scheduling, especially for the highly congested terminal-area airspace and for surface operations. GNSS will also play an important role in terrain-awareness warning systems.

New communication technologies, notably the automatic dependent surveillance-broadcast (ADS-B) service, will allow nearby aircraft to communicate their positions to each other as well as to air traffic controllers. Together, ADS-B and GNSS will be vital to enabling aircraft to detect and resolve conflicts, effectively decentralizing air traffic control responsibilities and allowing air traffic managers to handle increased traffic levels by focusing less on individual aircraft and more on the flow of groups of aircraft [22]. Automated conflict resolution will also be particularly critical in enabling and certifying the new generation of UAVs, allowing them to enter the airspace alongside conventional manned aircraft.

10.4.3 Backup Navigation Capabilities for Aviation

At one time, the GNSS was widely perceived as a cost-effective replacement for a patchwork of existing ground-based navigation aids, from ILS to TACAN, VOR, and DME beacons. A complete reliance on GNSS navigation, however, would introduce a significant vulnerability to RFI. Military GNSS users have long understood the need to protect GNSS signals against hostile jamming [2]. The vulnerability of civil GNSS applications to interference, intentional or unintentional, was first highlighted by a Volpe Center study in 2001 [23]. Instances of unintentional RFI have since been observed in the Monterey Bay, where faulty houseboat television receivers jammed GPS for several months in 2001 [24], and in regions of San Diego, where a Navy training exercise briefly jammed GPS in 2007 [25].

To maintain the availability of radio navigation in the case of a GNSS outage (due to RFI or due to other factors, such as scintillation during a major ionosphere storm), ground-based navigation aids must be employed. One possibility is that e-Loran will serve as a backup system for GNSS navigation for a wide range of applications, as discussed in Chapter 7. Another possibility is that a slimmed-down

network of navigation beacons will be maintained. In either case, it is hoped that some cost savings can be achieved by deactivating underutilized navigation aids.

10.5 Functionality of Aviation Augmentation Systems

During their early history, augmentation systems were perceived primarily as reference networks for generating and distributing differential GPS corrections. The role of augmentation systems in assuring navigation safety and reliability became more evident after a presidential order deactivated selective availability, which formerly dithered the GPS signal, thereby reducing navigation accuracy for civilian users by an order of magnitude. The subsequent improvement in the accuracy of stand-alone GPS might have diminished the role of GPS augmentations. However, these accuracy improvements made rare faults more salient. It became clear that augmentation systems would play a major role in verifying system integrity, given the risk posed by rare-event errors. GNSS satellites are essentially black boxes for aviation users and the general public. Since it is impractical for civil aviation authorities to certify GNSS *satellite hardware*, authorities must instead certify GNSS *signals*. To achieve this goal, aviation augmentation systems accomplish three major functions:

- Broadcast differential corrections to reduce navigation sensor error;
- Enable users to compute error bounds for on-line assessment of navigation performance;
- Generate prompt alert messages to warn users to exclude anomalous GNSS signals.

The remainder of this chapter develops an analytical basis for constructing safety guarantees based on these three functions, with a particular emphasis on the latter two capabilities, which are essential for all phases of flight. (By contrast the first capability, which improves accuracy through differential corrections, is not implemented in ABAS and is only necessary for precision approach and landing.)

10.5.1 Augmentation System Performance Requirements

The performance of an augmentation system in enabling safety-critical applications is quantified using a specific set of parameters that includes accuracy, integrity risk, continuity risk, and availability. Accuracy is the most familiar of these requirements, but other performance constraints are generally more difficult to satisfy in designing and certifying augmentation systems.

Accuracy is the nominal scatter of navigation errors. In GNSS applications, nominal errors typically follow an approximately Gaussian distribution. Accuracy is thus specified in terms of a Gaussian sigma value that describes the distribution's standard deviation.

Integrity risk is the probability with which a hazardously large navigation error occurs without the augmentation system issuing a timely alert. (This is the formal definition of HMI). Ideally, HMI would never happen. Unfortunately, the risk of

a large error can never be completely eliminated. In recognition of this practical reality, it is necessary to accept a finite (though very small) risk of HMI. As an example, the allowed signal-in-space integrity risk for LPV precision approach applications has been set at 10^{-7} per 150-second approach. For CAT-III precision landing operations, by contrast, the allowed integrity risk is 10^{-9} [4, 15]. Typically integrity risks are converted to a position-error bound called the *protection level*. Protection levels can be computed as a confidence interval on the PDF for the navigation error [15].

The duration within which a user must be warned of a significant navigation anomaly (one that introduces error levels that violate one or more protection levels) is called the time-to-alert (TTA). TTA accounts for the finite duration required for a monitor to detect an anomaly, for the augmentation system to relay the alert message to the aircraft, and for the aircraft to receive it (taking into account the possibility of one or more missed messages). Examples of system TTA requirements are 10s for LPV, 6s for category I, and 2s for category III.

Continuity risk is the likelihood that a monitor alert interrupts navigation. Monitor alerts due to detected failures cause the failed measurements to be excluded from use, an action that may in turn reduce navigation accuracy to a level at which the current flight operation is no longer safe. In addition, monitors may trigger false alerts. During precision operations, an alert that interrupts navigation service (whether real or false) is not merely a nuisance but a safety-critical event that may require drastic response (such as a go-around maneuver). Consequently for precision approach and landing operations, tight bounds are placed on the probability that a continuity break may occur (typically 10^{-5} or less per 15s). Continuity requirements in effect dictate monitor thresholds and determine monitor sensitivity. They also set maximum allowable failure probabilities for the underlying GNSS and, where relevant, the reference-station infrastructure that produces differential corrections.

Availability is the fraction of the time that integrity and continuity requirements can be satisfied by a navigation system. Because the geometry of the visible GNSS satellites continually changes, navigation system performance may not always satisfy operational requirements. To assess whether performance requirements are met, it is necessary to conduct an integrity test, which is generally implemented by verifying that the protection level (PL), which represents a worst-case error bound, does not exceed an operational maximum called the alert limit (AL). The availability test confirms that integrity (and potentially continuity and accuracy as well) is met for both nominal and faulted conditions.

$$\{PL \leq AL\} \rightarrow \text{System available} \quad (10.1)$$

To certify an aviation augmentation system requires an exhaustive validation that all of these safety requirements—accuracy, integrity, continuity, and availability—are satisfied.

10.5.2 Error Bounding Under Nominal Conditions

For users to assess availability on the fly, augmentation systems must enable computation of a navigation system error bound in the form of a PL. Availability is

highest for systems with tighter error bounds (and smaller PLs). PLs for ABAS are larger than those for other augmentation systems that leverage differential corrections to improve accuracy. Regional differential corrections help tighten error bounds for SBAS and GRAS. Even tighter PLs are possible for GBAS, which leverages precise local-area corrections.

Typically, distinct PLs are evaluated in the vertical and horizontal direction. Because GPS is more accurate in the horizontal direction than the vertical (because satellites are visible in the hemisphere above the user and not below) and because the consequences of vertical errors are particularly severe, the vertical PL typically exceeds the horizontal PL as the primary limiter of system availability.

For a given integrity risk specification, the vertical protection level (VPL) can be computed as a confidence limit proportional to the standard deviation of the vertical positioning error σ_v .

$$VPL = K\sigma_v \quad (10.2)$$

If the error distribution is modeled as Gaussian, the value of the proportionality constant may be computed by inverting the Gaussian cumulative distribution function (e.g. $K = 5.85$ for a two-sided integrity risk of $5 \cdot 10^{-9}$).

PLs for different users, even nearby users, are not necessarily the same, as individual users may navigate with different subsets of the visible GNSS satellites. Similarly, differential position corrections computed using SBAS, GRAS, or GBAS may differ among users of a given system. To support arbitrary satellite subsets, augmentation systems broadcast error-bounding parameters and differential corrections for each satellite separately. (In the case of SBAS, separate corrections for ionosphere delay are also broadcast based on a regional grid [27], as described in Chapter 4.) User avionics compute PLs using the error-distribution parameters broadcast for each visible satellite. For a subset of N satellites, the total vertical error E_v is the sum of the satellite ranging error for each satellite E_i projected into the vertical direction and weighted by its relative accuracy (through a set of linearized navigation-solution coefficients s_i).

$$E_v = \sum_{i=1}^N s_i E_i \quad (10.3)$$

Assuming unbiased, uncorrelated Gaussian error distributions for each satellite, the variances of the weighted satellite errors add to give the variance of the vertical positioning error. Thus, if each satellite-ranging error is characterized by a standard deviation σ_i , the standard deviation of the vertical position error is the following.

$$\sigma_v = \sqrt{\sum_{i=1}^N (s_i \sigma_i)^2} \quad (10.4)$$

Equation (10.4) is frequently used to convert range-domain error bounds into a position-domain error bound because (10.2) and (10.4) can be computed quickly on-the-fly and because error distributions for each satellite are described by a

single parameter, σ_i , which can easily be transmitted over a finite-bandwidth communication link (e.g., in SBAS, GBAS, or GRAS). By comparison, a more sophisticated vertical-error analysis using a nonparametrically defined joint probability distribution function for satellite errors is not yet practical for aviation applications.

Although (10.4) is convenient to implement, the idealized assumptions used in constructing this equation are not necessarily valid. This is to say that actual error distributions are rarely unbiased, uncorrelated, and Gaussian. A biased distribution may result, for instance, from phase-center calibration errors, which introduce an offset between the calibrated and actual locations of a GNSS antenna. Correlation errors may result from similarities among ionosphere delays for satellites in the same region of the sky [28], especially for single-frequency applications with regional differential corrections (like SBAS and GRAS) [29] or with no differential corrections at all (like ABAS). Non-Gaussian error distributions are also common. In particular, intermittent multipath may result in error distributions with non-Gaussian features, such as asymmetry or heavy tails [30].

In order to certify augmentation system integrity given nonideal navigation error distributions (which may be biased, correlated, and non-Gaussian), great care must be taken to ensure that the VPL calculated using (10.2) and (10.4) is valid. To this end, the aviation community has developed procedures for replacing actual error distributions with conservative Gaussian models (which *are* unbiased, uncorrelated, and Gaussian). These conservative models are often called *overbounds*. The critical property of an overbound is that it can be used to obtain a confidence interval (such as a PL) that is guaranteed to be at least as wide as the confidence interval for the actual error distribution. Wider protection levels are conservative in the sense that they are always unavailable when integrity criteria are not met. However, overbounds that significantly inflate the actual protection level will degrade system availability. To maximize system availability (while always ensuring integrity), it is desired that overbounds be as “tight” as possible, resulting in a modeled VPL only slightly larger than the actual VPL.

For GNSS navigation, a major challenge has been to define overbounds that apply to linear combinations of random variables since position errors are linear combinations (i.e., weighted sums) of ranging source errors [see (10.3)]. A method for bounding individual ranging source errors in order to construct a conservative bound for the vertical positioning error was first identified by DeCleene [31]. DeCleene noted that a conservative protection level could be computed from a set of range-domain overbounds so long as the integrated probability in the tails of each “overbounding” model was greater than the integrated probability in the tails of the corresponding actual distribution. In other words the integrated overbound G_o must be greater than the cumulative distribution function (CDF) for the actual error (G_a) below the median and less than the actual CDF above the median.

$$\begin{aligned} G_o(x) &\geq G_a(x), \quad \forall G_a < \frac{1}{2} \\ G_o(x) &\leq G_a(x), \quad \forall G_a \geq \frac{1}{2} \end{aligned} \tag{10.5}$$

If the actual distribution is Gaussian and zero-mean, for example, any Gaussian distribution with a standard deviation larger than the actual value is an overbound. Figure 10.3 illustrates a non-Gaussian distribution and a Gaussian overbound that satisfies the conditions of (10.5).

Unfortunately, DeCleene's original overbounding formulation does not apply to arbitrary distributions. Rather, the derivation requires that the actual distributions be independent, symmetric, and unimodal (single-peaked). To address this limitation, other overbounding methods have been introduced to generalize the original concept. These alternative overbounding methods include techniques for bounding arbitrary-shaped, independent distributions [32–34] and for bounding symmetric, correlated distributions [35]. Ultimately, the major limitation of all of these methods is the fact that statistical sampling methods can never fully characterize an actual error distribution at the extreme values in the tails. Ultimately, engineering judgment is necessary to analyze nonstationary data sets [36] and to apply extreme value theory or other rational methodologies for extrapolating the far tails of the actual error distribution [37].

In practice, appropriate Gaussian models can usually be defined that overbound the available statistical data representing actual ranging measurement errors. Sigma parameters for these Gaussian models can be tabulated a priori [38]. As an example, GBAS users compute sigma values for each satellite based on overbound models for each of four major error sources: ground receiver errors ($\sigma_{gnd,i}$), airborne receiver errors ($\sigma_{air,i}$), and spatial decorrelation errors associated with ionosphere ($\sigma_{iono,i}$) and troposphere ($\sigma_{tropo,i}$) gradients.

$$\sigma_i = \sqrt{\sigma_{gnd,i}^2 + \sigma_{air,i}^2 + \sigma_{iono,i}^2 + \sigma_{tropo,i}^2} \quad (10.6)$$

In GBAS, $\sigma_{air,i}$ is different for each aircraft and must be stored in memory in the user avionics (as a function of satellite elevation). Other sigma values are

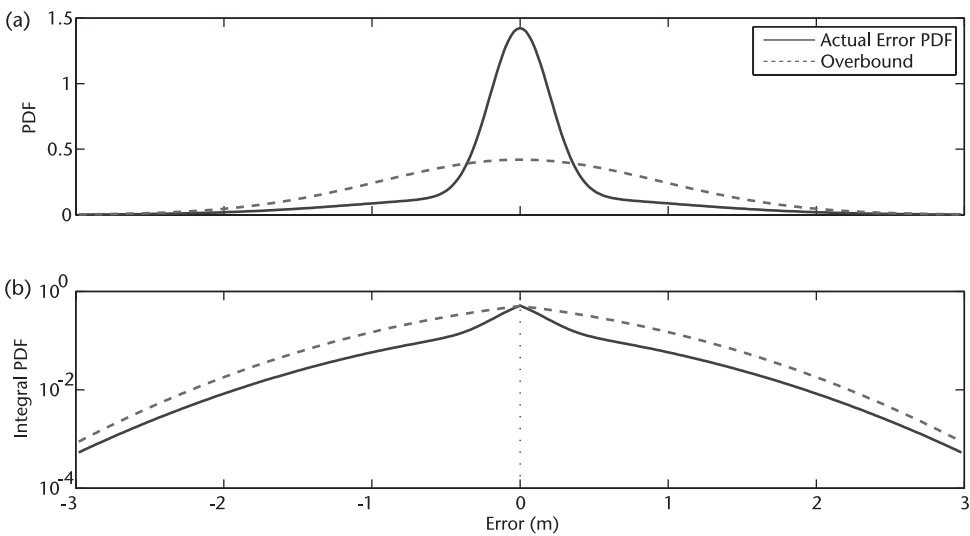


Figure 10.3 Gaussian overbound for a non-Gaussian error distribution: (a) PDFs for the overbound and actual distribution and (b) integrated probability (from $\pm\infty$ to zero) for each distribution.

constructed by the user avionics based on parameters broadcast by the GBAS ground facility [39]. The values of these parameters are all established off-line during GBAS design and certification.

Limited communication bandwidth restricts the resolution of broadcast parameters. For GBAS, sufficient bandwidth is available through the VHF data broadcast to represent ground-receiver errors as floating point numbers [39]. For SBAS, by contrast, sigma values are represented more coarsely. GEO correction messages are limited to only 250 bps, so SBAS sigma values are represented with a 16-level (4-bit) discretization [40].

10.5.3 Error Bounding Under Anomalous Conditions

Protection levels must bound both nominal and anomalous conditions. Protection levels for anomalous conditions ensure integrity when a monitor fails to detect a navigation fault in a timely manner. Examples of anomalous conditions include GNSS satellite faults (such as a clock failure), atmospheric anomalies (such as an ionosphere storm), or augmentation system faults (such as a receiver or datalink failure). Chapter 4 provides a more extensive treatment of relevant fault modes (for uncorrected and for differentially corrected GNSS navigation).

10.5.3.1 Fault-Mode Protection Levels

Because signal monitors may be slow to detect (or fail to detect) an anomaly, and because these monitors can only reliably detect anomalies above a certain magnitude, protection levels are needed to ensure integrity during anomalous events. For most augmentation systems (SBAS, GBAS, GRAS), fault modes are identified by a FMEA. Based on the failure modes and effective analysis (FMEA), individual PLs can be derived for each faulted condition ($PL_{fault,k}$) to accompany the PL for nominal conditions (PL_{nom}). The overall PL, used to determine system availability, is set equal to the largest of these individual error bounds.

$$PL = \max\left(\max_{\text{all faults}} (PL_{fault,k}), PL_{nom}\right) \quad (10.7)$$

This process of creating distinct PLs for each fault mode simplifies integrity analysis. The combined error distribution, covering both nominal and faulted conditions, is generally non-Gaussian and difficult to model [41]. In particular, uncertain fault-mode parameters complicate the process of defining a single, overall error distribution for all conditions. Decomposing integrity analysis for distinct faulted and nominal conditions, by contrast, makes Gaussian overbounding assumptions more reasonable (and the resulting error bounds tighter). Decomposition also permits more rigorous bounding of uncertain model parameters.

Within the aviation community, a methodology for deriving fault-mode error bounds has been developed based on the following assumptions.

- Faults occur rarely. The frequency of occurrence of any type of fault is thus assumed to be less than or equal to a specified prior probability P_{fault} [42].

- Faults may impact either one satellite (e.g., a satellite clock failure) or multiple satellites at a time (e.g., a severe ionosphere storm).
- Errors for faulty measurements can be decomposed into two parts: a random error and a fault-induced bias. The random-error PDF is assumed to be equivalent to that for nominal conditions. The fault-mode bias is assumed to take an arbitrary value somewhere within a threat space defined by the FMEA effort.
- Nominal errors levels are assumed for all measurements not directly impacted by the anomaly.

Individual fault-mode PLs can be derived based on these assumptions. The total integrity risk budget must first be divided among the PLs for nominal and all faulted conditions. As a hypothetical example, 10 PLs might be defined by evenly dividing an integrity risk budget of 10^{-7} among nine fault modes and the nominal condition, allowing an integrity risk suballocation of 10^{-8} for each of the 10 PLs. A fault tree for this simple example is illustrated in Figure 10.4. More highly detailed fault trees for certified augmentations may be found in the relevant HMI validation documents for those systems [29]. Although integrity allocations could conceptually be assigned in a dynamic, time-varying manner to optimize system performance, VPL values are remarkably insensitive to large variations in the integrity risk allocations because they are derived from Gaussian error models. Consequently, static allocations are used in current-generation augmentations.

In essence, each PL is simply a confidence interval. This interval is defined to bound all but a small fraction of errors (whose integrated probability does not exceed the integrity risk suballocation). The PL for a generic fault can be easily derived for the case of Gaussian errors. In the vertical direction, the PL for the generic fault includes one term proportional to the nominal standard deviation for the vertical error (σ_v) and a second term that describes the fault-mode bias (μ). The sign of the fault-mode bias is assumed to be unknown.

$$VPL_{fault} = K_{fault} \sigma_v + \mu \quad (10.8)$$

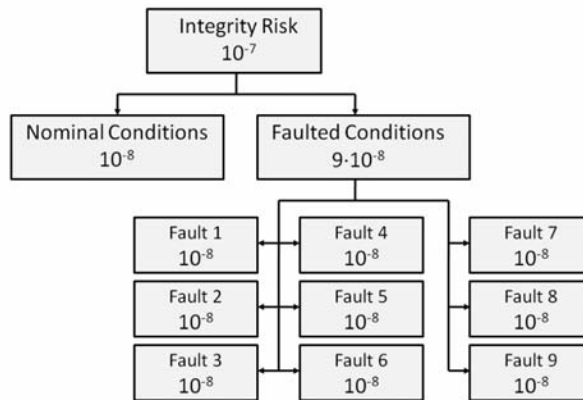


Figure 10.4 Fault tree example: Integrity risk budget suballocated to nominal and faulted conditions.

The scaling parameter K_{fault} is computed to bracket all but a small fraction of large errors. This small probability is labeled P_{bnd} (and is equal to the probability that the vertical position error exceeds the VPL given that a fault has occurred and that monitors have missed detecting it). K_{fault} is derived from P_{bnd} by inverting the zero-mean Gaussian cumulative distribution Q .

$$K_{fault} = -Q^{-1}(P_{bnd}) \quad (10.9)$$

The probability P_{bnd} can be written in terms of the integrity risk allotment (P_a) suballocated from the overall integrity risk budget, the probability that the fault occurs (P_{fault}), and the probability of monitor missed detection given faulted conditions (P_{md}).

$$P_{bnd} = \frac{P_a}{P_{md} P_{fault}} \quad (10.10)$$

Fault-mode PLs are not always evaluated in the user avionics. The following two examples of fault-mode PLs include one case explicitly evaluated by users in real time (GBAS reference-receiver-fault VPL) and a second case implicitly evaluated off-line as part of the certification process (satellite-clock excessive acceleration VPL).

10.5.3.2 Example: GBAS Reference-Receiver Fault

An example from GBAS is the fault-mode PL for receiver equipment failures. GBAS requires users to compute a VPL that bounds the vertical position error when one of the GBAS ground facility's reference receivers is faulty, a case designated as "hypothesis 1" or H1. During an H1 event, monitors have not yet detected the receiver fault. The following equation can be used to evaluate the VPL for the H1 event involving a fault-induced bias B_j that affects a particular reference receiver (identified by the index j) [43].

$$VPL_{H1,j} = K_{H1} \sigma_v + B_j \quad (10.11)$$

The parameter K_{H1} is computed using (10.9) and (10.10). For example, a category-I GBAS system has a total integrity risk budget of $2 \cdot 10^{-7}$. The fraction allocated to the VPL for the H1 fault ($P_{a,H1}$) is defined to be $2 \cdot 10^{-8}$ for a typical GBAS ground facility with four reference receivers. The probability of an undetected receiver fault ($P_{md,H1} \cdot P_{fault,H1}$) must not exceed $1 \cdot 10^{-5}$, so K_{H1} is equal to 2.88 [44].

$$K_{H1} = -Q^{-1}\left(\frac{2 \cdot 10^{-8}}{10^{-5}}\right) = 2.88 \quad (10.12)$$

In the general case, any of the ground receivers might be affected. Thus, a distinct $VPL_{H1,j}$ is computed for each reference receiver j . The bias is estimated

as a weighted sum of the so-called B-values for each satellite and receiver $B_{i,j}$, which are broadcast from the ground station to users.

$$B_j = \left| \sum_i s_i B_{i,j} \right| \quad (10.13)$$

The ground station computes each B-value by differencing the pseudorange correction for a particular satellite i (averaged over all receivers) from a modified pseudorange correction averaged over all receivers except the hypothetically faulty receiver, which is identified by the index j [26].

10.5.3.3 Example: Excessive Acceleration Fault

To simplify broadcast messages and user avionics, most fault-mode PLs are not explicitly evaluated by the user avionics. Instead, integrity can be validated by an off-line analysis showing that a particular fault-mode PL can never exceed the PL expression that the user evaluates in real time. Note that off-line analysis may also indicate that the explicitly evaluated PL does not bound all unevaluated fault mode PLs for all visible satellite geometries. In this case, broadcast sigma values must be increased enough to ensure a bound. This process is called *sigma inflation*. The precise level of sigma inflation required to bound the PL for a particular fault mode can be determined either by an analytical range-domain analysis [45–47] or by extensive simulations [48].

An example of a fault-mode PL not explicitly evaluated by GBAS users is the one that models an excessive-acceleration (EA) fault, in which a satellite's clock suddenly accelerates away from the time reference used by the other satellites in the GPS constellation. The VPL for an EA event has the following form.

$$VPL_{EA} = K_{EA} \sigma_v + \mu_{EA} \quad (10.14)$$

The fault-mode bias impacts only one satellite. However, as any of the i satellites used in the navigation solution are at risk for a clock acceleration event, the VPL is conservatively computed assuming the worst satellite is affected by the EA ranging error, e_{EA} .

$$\mu_{EA} = \max_i |s_i e_{EA}| \quad (10.15)$$

The EA threat model consists of a single parameter (clock acceleration), which impacts both the fault-mode error μ_{EA} and the probability of missed detection $P_{md,EA}$. The worst threat is not necessarily a large acceleration, since monitors are more likely to detect larger acceleration values (resulting in reduced $P_{md,EA}$ and K_{EA}) even though the size of the measurement error increases (larger μ_{EA}). The relationship between μ_{EA} and $P_{md,EA}$ for all threats in the EA threat space can be determined through simulation (Figure 10.5). For any particular satellite geometry, a worst-case threat can be identified that maximizes VPL_{EA} computed using (10.14) [49]. The level of sigma inflation is determined to ensure that the

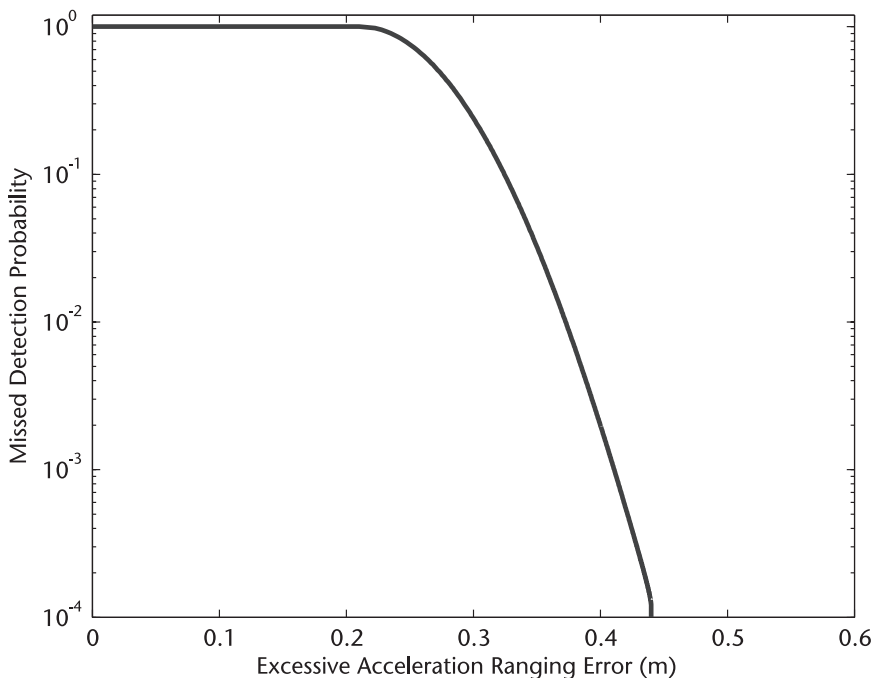


Figure 10.5 Relationship between missed-detection probability ($P_{md,EA}$) and ranging bias (μ_{EA}) for the excess acceleration fault.

user VPL expression always exceeds the worst-case VPL_{EA} (for all relevant satellite geometries).

10.5.4 Monitoring

GNSS faults, though rare, may result in errors of tens or hundreds of meters in magnitude. These massive errors introduce unacceptable risks for most aviation users, except possibly those in the *en route* or oceanic phases of flight. Because fault-mode errors can be so large, it is critical not just to bound anomalous measurements but also to exclude them from the navigation solution. Otherwise, PLs would be too large to support the ALs associated with precision aviation applications, which vary in magnitude from 10m for category-III landing to 50m for an LPV approach.

Integrity monitors are the mechanism by which augmentation systems isolate and exclude anomalous measurements. Integrity monitors operate by comparing a monitor statistic to a threshold; if the statistic ever exceeds the threshold, all related GNSS measurements are flagged as suspicious. Further monitor logic may subsequently process suspicious measurements to identify a particular receiver, a particular frequency, a particular satellite, or a particular satellite constellation as faulted. If a fault is declared, the augmentation system then warns users not to navigate using any measurements associated with the fault. Eventually, a strict readmittance procedure may allow the faulted measurements to be declared healthy and once again be used for navigation [50].

Different types of monitors are used for different applications. ABAS, for example, relies on general-purpose monitors that can detect a wide range of threats. These monitors are less sensitive, but somewhat easier to implement, than the specialized monitors used in other augmentation systems (SBAS, GBAS, GRAS) to detect specific fault modes identified by an FMEA. Examples of both monitor classes (an ABAS residual monitor and a GBAS monitor for the EA threat) are described in more detail in the following.

10.5.4.1 Example: ABAS Residual Monitor

Typical ABAS implementations rely on general-purpose monitors based on the residuals of the navigation solution [51]. The navigation solution determines position and time using an iterative least-squares solution. For each iteration n , corrections to the position and time vector $\delta \mathbf{p}_{n+1}$ are computed using the weighted projection matrix \mathbf{S}_n and a residual vector \mathbf{r}_n from the previous iteration. The residual is a difference between the measured pseudoranges $\boldsymbol{\rho}_{meas}$ and the computed pseudoranges $\boldsymbol{\rho}_n$ based on the current best estimate of user position and time.

$$\delta \mathbf{p}_{n+1} = \mathbf{S}_n \mathbf{r}_n = \mathbf{S}_n (\boldsymbol{\rho}_{meas} - \boldsymbol{\rho}_n) \quad (10.16)$$

When the position solution converges (that is, when the entries of $\delta \mathbf{p}_{n+1}$ approach zero), a nonzero residual vector remains. (This converged residual vector lies in the null space of the matrix \mathbf{S}_n). Outliers can be identified by comparing each element of the residual vector to a threshold and excluding any measurements that exceed the threshold. This approach can be used to detect a wide range of errors; however, the method has trouble detecting small errors (because pseudorange measurements used to compute residuals are somewhat noisy) and may fail to detect errors involving multiple faulted satellites (unless the fault vector lies in the null space of \mathbf{S}_n).

10.5.4.2 Example: GBAS Excessive Acceleration Monitor

To improve monitor sensitivity and to ensure observability of faults involving more than one satellite, augmentation systems with a terrestrial infrastructure (SBAS, GBAS, GRAS) use specialized monitors to detect specific faults identified by their respective FMEAs. A representative integrity monitoring architecture is illustrated in Figure 10.6. In this architecture, the results from multiple specialized monitors feed into an executive monitor (EXM) which ultimately determines whether or not a particular satellite or reference receiver should be declared unhealthy.

An example of a specialized monitor is the EA monitor implemented in GBAS. This monitor is designed to detect satellite clock acceleration anomalies (as described in Section 10.5.3). To detect EA faults with high sensitivity, precise carrier-phase measurements for each satellite are processed separately. For each satellite i , the monitor statistic $m_{EA,i}$ estimates excess acceleration as the second derivative, or *acceleration*, of the carrier-phase ranging measurements ϕ_i . In (10.17), the second-derivative operator can be represented by any discrete approximation that cancels out the carrier-phase integer ambiguity.

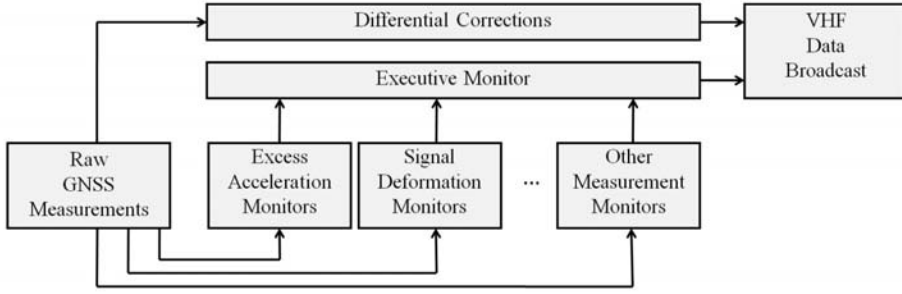


Figure 10.6 Integrity monitoring architecture for GBAS.

$$m_{EA,i} = \frac{d^2}{dt^2}(\phi_i) - a_{i,nom} \quad (10.17)$$

To ensure that the monitor statistic is unbiased, a nominal level of expected pseudorange acceleration, $a_{i,nom}$, is removed from the excessive acceleration estimate. For the case of a fixed reference antenna, this nominal acceleration can be attributed purely to typical satellite clock variation and to satellite motion, characterized through the broadcast ephemeris data.

The magnitude of the threshold for the EA monitor and the noise level of the monitor statistic determine the likelihood that the monitor detects an acceleration fault. Hence, these parameters also determine the missed-detection probability $P_{md,EA}$ [which defines K_{EA} in (10.14)]. Specifically, $P_{md,EA}$ can be computed by integrating the probability that the noisy monitor statistic lies below the threshold (given that the monitor statistic distribution is biased by the fault). This procedure for computing the missed detection probability is illustrated in Figure 10.7(b). Different values of $P_{md,EA}$ (and μ_{EA}) are associated with each fault in the EA threat space.

A tighter threshold provides greater monitor sensitivity and, hence, a lower value of $P_{md,EA}$. However, the magnitude of the threshold is limited by the specification for continuity risk. More specifically, the threshold determines the rate of monitor false alarms (which occur when the test statistic exceeds the threshold even though no fault is present). Typically, and very conservatively, false alarms are assumed to always result in continuity breaks that interrupt navigation. Hence, the continuity risk budget can be subdivided among all monitors and all measurements being monitored to determine an allowable false alarm rate for each monitor. The threshold is set so that the integral of the measurement noise distribution beyond the threshold (for the nominal, unbiased case) does not exceed the continuity risk allocation to false alarms. This procedure for computing the threshold is illustrated in Figure 10.7(a).

A slight refinement accounts for one additional design requirement: TTA. The TTA requirement dictates the time available for monitors to alert users that a severe anomaly has occurred (that is, an anomaly not bounded by the user PL). The probability of missed detection must be low enough to meet integrity require-

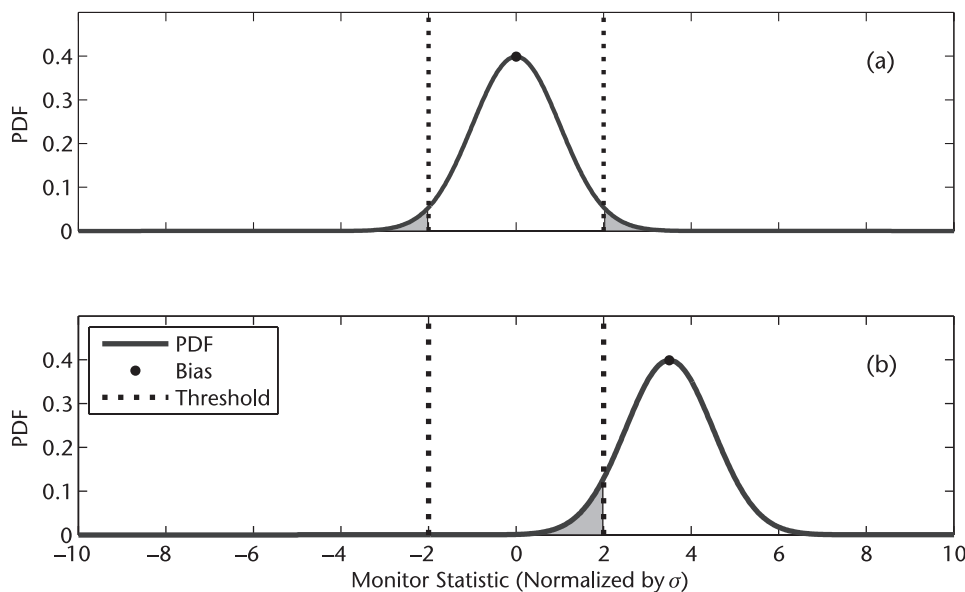


Figure 10.7 Quantifying monitor performance in terms of continuity and integrity. (a) Continuity: the false alarm rate is the integral of the nominal error distribution outside the threshold. (b) Integrity: the missed detection probability is the integral of the fault-case error distribution inside the threshold.

ments (for an error bias μ) within the TTA after the user first experiences that error. It is possible to account for TTA in developing the curve relating error magnitude μ to the monitor probability of missed detection P_{md} , as illustrated in Figure 10.5. In effect, the TTA shifts the time references for the missed-detection risk (P_{md}) transient and for the error bias (μ) transient, slightly modifying the error characteristic shown in Figure 10.5 [47].

10.6 Conclusion

For aviation applications of GNSS, the overriding concern is validation of system safety. To ensure safety of GNSS navigation, as quantified in terms of accuracy, integrity, continuity, and availability, aviators will rely on augmentation systems to monitor and bound GNSS errors. This chapter outlines a mechanism for ensuring integrity and continuity by comparing protection levels to alert limits in real time. Achieving tight bounds that result in high availability requires the use of monitors that exclude rare system faults. By analyzing the time transients for monitor missed-detection probability and shifting by the TTA, fault-mode protection levels can be constructed. Typically, fault-mode protection levels are smaller than the protection level for nominal conditions and need not be evaluated on-the-fly by user avionics.

Aviation users have benefited enormously from GNSS as a tool to enable worldwide area navigation. A backup navigation infrastructure (consisting of a depleted set of VOR, DME, and TACAN stations as well as LORAN) will remain essential, as GNSS signals are vulnerable to intentional or unintentional radio frequency interference. Nonetheless, new precision landing and air traffic control

applications will make GNSS an indispensable component of the aviation infrastructure in the 21st century.

10.7 Further Reading

An insightful discussion of GNSS modernization and its impact on aviation applications can be found in Hegarty and Chartre [4]. More information regarding ongoing developments of the SBAS may be found in Walter et al [19] and of the GBAS in Murphy et al [1]. The connections between monitoring, fault-mode bounding, and TTA requirements are more fully developed in [47, 49].

References:

- [1] Murphy, T., and T. Imrich, "Implementation and Operational Use of (GBAS) Ground Based Augmentation Systems—A Component of the Future Air Traffic Management System," *Proceedings of the IEEE*, Vol. 96, No. 12, December 2008, pp. 1936–1957.
- [2] Rife, J., et al, "Navigation, Interference Suppression, and Fault Monitoring in the Sea-Based Joint Precision Approach and Landing System," *Proceedings of the IEEE*, Vol. 96, No. 12, December 2008, pp. 1958–1975.
- [3] Crosby, G., et al, "A Ground-Based Regional Augmentation System (GRAS)—The Australian Proposal," *Proc. ION-GPS 2000*, Salt Lake City, UT, Sept. 19–22, 2000.
- [4] Hegarty, C., and E. Chartre, "Evolution of the Global Navigation Satellite System (GNSS)," *Proceedings of the IEEE*, Vol. 96, No. 12, December 2008, pp. 1902–1917.
- [5] Dodington, S., "Radio Navigation," *Avionics Navigation Systems*, M. Kayton and W. Fried, eds., New York, NY: John Wiley & Sons, 1969.
- [6] Braff, R., J. D. Powell, and J. Dorfler, "Applications of the GPS to Air Traffic Control," *Global Positioning System: Theory and Applications*, Vol. II, Progress in Astronautics and Aeronautics Series, Vol. 164, B. Parkinson, and J. Spilker, eds., Washington, D.C.: AIAA 1996.
- [7] Eschenbach, R., "GPS Applications in General Aviation," *Global Positioning System: Theory and Applications*, Vol. II, Progress in Astronautics and Aeronautics Series, Vol. 164, B. Parkinson, and J. Spilker, eds., Washington, D.C.: AIAA 1996.
- [8] U.S. Department of Defense and Department of Transportation, *2001 Federal Radionavigation Systems*, <http://www.navcen.uscg.gov/pubs/frp2001/FRS2001.pdf>.
- [9] Siouris, G. M., *Aerospace Avionics Systems: A Modern Synthesis*, San Diego, CA: Academic Press, Inc., 1993.
- [10] Kim, E.-H., T. Walter, T., and J. D. Powell, "WAAS-Based Flight Inspection System," *Proc. ION Annual Meeting*, Cambridge, MA, April 23–25, 2007.
- [11] U.S. Department of Defense, Department of Homeland Security, and Department of Transportation, *2005 Federal Radionavigation Plan*, <http://www.navcen.uscg.gov/pubs/frp2005/2005%20FRP%20WEB.pdf>.
- [12] Davis, M., "What's Next for WAAS?," *SatNav News*, Vol. 33, March 2008, pp. 1–2. http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/gnss/library/satNav/media/SatNav_March08.pdf.
- [13] Luo, M., et al, "Ionosphere Threat to LAAS: Updated Model, User Impact, and Mitigations," *Proc. of ION-GNSS*, Long Beach, CA, Sept. 21–24, 2004. pp. 2771–2785.
- [14] Datta-Barua, S., et al, "Using WAAS Ionospheric Data to Estimate LAAS Short-Baseline Gradients," *Proc. ION National Technical Meeting*, San Diego, CA, Jan. 28–30, 2002. pp. 523–530.

- [15] Pullen, S., T. Walter, and P. Enge, P., "System Overview, Recent Developments, and Future Outlook for WAAS and LAAS," *Tokyo Univ. of Mercantile Marine GPS Symposium*, Tokyo, Japan, Nov. 11–13, 2002. <http://waas.stanford.edu/~www/papers/gps/PDF/PullenTokyo02.pdf>.
- [16] Misra, P., and P. Enge, *Global Position System: Signals, Measurements, and Performance*, Lincoln, MA: Ganga-Jamuna Press, 2006.
- [17] Lee, Y., "Receiver Autonomous Integrity Monitoring Availability for GPS Augmented with Barometric Altimeter Aiding and Clock Coasting," *Global Positioning System: Theory and Applications*, Vol. II, Progress in Astronautics and Aeronautics Series, Vol. 164, B. Parkinson and J. Spilker, eds., Washington, D.C.: AIAA, 1996.
- [18] Hwang, P., G. McGraw, and J. Bader, "Enhanced Differential GPS Carrier-Smoothed Code Processing Using Dual-Frequency Measurements," *NAVIGATION: Journal of the ION*, Vol. 46, No. 2, Summer 1999, pp. 127–138.
- [19] Walter, T., et al, "Worldwide Vertical Guidance of Aircraft Based on Modernized GPS and New Integrity Augmentations," *Proceedings of the IEEE*, Vol. 96, No. 12, December 2008, pp. 1918–1935.
- [20] Joint Program and Development Office (JPDO). *Integrated Work Plan for the Next Generation Air Transportation System Version 1.0*. 2008. <http://www.jpdo.gov/iwp.asp>.
- [21] Graham, R., et al, "Absolute Versus Relative Navigation: Theoretical Considerations from an ATM Perspective," *FAA/Eurocontrol Air Traffic Management R&D Seminar*, Budapest, Hungary, June 2003.
- [22] Sridhar, B., S. Grabbe, and A. Mukherjee, "Modeling and Optimization in Traffic Flow Management," *Proceedings of the IEEE*, Vol. 96, No. 12, December 2008, pp. 2060–2080.
- [23] Volpe National Transportation System Center, *Vulnerability assessment of the transportation infrastructure relying on the Global Positioning System*, Washington, D.C., 2001.
- [24] Clynch, J., et al, "Multiple GPS RFI Sources in a Small California Harbor," *Proc. of ION-GPS*, Portland, OR, Sept. 24–27, 2002.
- [25] Jewel, D., "GPS Insights: JNC Briefing on Jamming Incident," *GPS World*, April 2007. <http://mg.gpsworld.com/gpsmg/article/articleDetail.jsp?id=418420&searchString=april%202007>.
- [26] Braff, R., "Description of the FAA's local area augmentation system (LAAS)," *NAVIGATION: Journal of the ION*, Vol. 44, No. 4, 1997, pp. 411–423.
- [27] Walter, T., "WAAS MOPS: Practical Examples," *Proc. ION National Technical Meeting*, San Diego, CA, Jan 25–27, 1999.
- [28] Walter, T., P. Enge., and A. Hansen, "An Integrity Equation for Use with Space Based Augmentation Systems," *Proc. GNSS-97*, Munich, Germany, April 1997.
- [29] Raytheon, *Algorithm Contribution to HMI for the Wide Area Augmentation System*, Raytheon Company unpublished Work, 2002.
- [30] Braff, R., and C. Shively, "A Method of Over Bounding Ground Based Augmentation System (GBAS) Heavy Tail Error Distributions," *Journal of Navigation*, Vol. 52, No. 2, 2005, pp. 83–103.
- [31] DeCleene, B., "Defining Pseudorange Integrity—Overbounding," *Proc. ION-GPS*, Salt Lake City, UT, Sept. 19–22, 2000.
- [32] Rife, J., S. Pullen, and B. Pervan, "Core Overbounding and Its Implications for LAAS Integrity," *Proc. ION-GNSS*, Long Beach, CA, Sept. 21–24, 2004.
- [33] Rife, J., T. Walter, and J. Blanch, "Overbounding SBAS and GBAS Error Distributions with Excess-Mass Functions," *Proc. IGNS*, Sydney, Australia, Dec. 6–8, 2004.
- [34] Rife, J., et al, "Paired Overbounding for Nonideal LAAS and WAAS Error Distributions," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 42, No. 4, 2006, pp. 1386–1395.
- [35] Rife, J., and D. Gebre-Egziabher, "Symmetric Overbounding of Correlated Errors," *NAVIGATION: Journal of the ION*, Vol. 54, No. 2, 2007, pp. 109–124.

- [36] Pervan, B., and I. Sayim, "Sigma Inflation for the Local Area Augmentation of GPS," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 37, No. 4, 2001, pp. 1301–1311.
- [37] Shively, C., and R. Braff, "An Overbound Concept For Pseudorange Error from the LAAS Ground Facility," *Proc. ION National Technical Meeting*, San Diego, CA, June 26–28, 2000.
- [38] McGraw, G., et al, "Development of LAAS Accuracy Models," *Proc. ION-GPS*, Salt Lake City, UT, Sept. 19–22, 2000.
- [39] RTCA Inc., *GNSS-Based Precision Approach Local Area Augmentation System (LAAS) Signal-in-Space Interface Control Document (ICD)*, Washington, D.C.: RTCA/DO-246B, 2001.
- [40] RTCA Inc., *Minimum Operational Performance Standards for GPS/Wide Area Augmentation System Airborne Equipment*, Washington, D.C.: RTCA/DO-229D, 2006.
- [41] Schroth, G., et al, "Failure Detection and Exclusion via Range Consensus," *Proc. European Navigation Conference*, Toulouse, France, April 23–25, 2008.
- [42] Pullen, S., J. Rife, and P. Enge, "Prior Probability Model Development to Support System Safety Verification in the Presence of Anomalies," *Proc. of IEEE/ION PLANS*, San Diego, CA, April 25–27, 2006.
- [43] Liu, F., "Development of Gaussian Overbounds for the LAAS Signal-in-Space Integrity Monitoring Algorithms," *NAVIGATION: Journal of the ION*, Vol. 46, No. 1, 1999, pp. 49–64.
- [44] RTCA Inc., *Minimum Aviation System Performance Standards for Local Area Augmentation System (LAAS)*, Washington, D.C.: RTCA/DO-245A, 2004.
- [45] Shively, C., "Derivation of Acceptable Error Limits for Satellite Signal Faults in LAAS," *Proc. ION-GPS*, Nashville, TN, Sept. 14–17, 1999.
- [46] Zaugg, T., "A New Evaluation of Maximum Allowable Errors and Missed Detection Probabilities for LAAS Ranging Source Monitors," *Proc. ION Annual Meeting*, Albuquerque, NM, June 24–26, 2002.
- [47] Rife, J., and R. E. Phelts, "Formulation of a Time-Varying Maximum Allowable Error for Ground-Based Augmentation Systems," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 44, No. 2, 2008, pp. 548–560.
- [48] Lee, J., et al, "Position-Domain Geometry Screening to Maximize LAAS Availability in the Presence of Ionosphere Anomalies," *Proc. ION-GNSS*, Fort Worth, TX, Sept. 22–25, 2006.
- [49] Rife, J., S. Pullen, and P. Enge, "Evaluating Fault-Mode Protection Levels at the Aircraft in Category III LAAS," *Proc. ION Annual Meeting*, Cambridge, MA, April 23–25, 2007.
- [50] Xie, G., *Optimal On-Airport Monitoring of the Integrity of GPS-Based Landing Systems*, Ph.D. thesis, Stanford University Department of Aeronautics and Astronautics, Stanford, CA, 2004. <http://waas.stanford.edu/~www/papers/gps/PDF/Thesis/GangXieThesis04.pdf>.
- [51] Lee, Y., "Performance of Receiver Autonomous Integrity Monitoring (RAIM) in the Presence of Simultaneous Multiple Satellite Faults," *Proc. ION Annual Meeting*, Dayton, OH, June 7–9, 2004.

Integrated GNSS and Loran Systems

Sherman Lo and Ben Peterson

11.1 Introduction

This chapter examines the integration of GNSS with long-range navigation (Loran), a terrestrial, pulsed, low-frequency (LF), horizontal navigation system [1, 2]. This integrated system is primarily aimed at applications where safety-of-life is paramount such as those described in Chapter 10. These are applications requiring a high-level of integrity. Loran, as its name suggests, uses signals that propagate over a large distance. Due to the nature of LF propagation and the power of the Loran transmission, users at distances of 800 km or more can receive these signals enabling the long-range navigation system. The susceptibility of GNSSs to service interruptions from unintentional or malicious RFI has spurred interest in identifying and developing suitable GNSS backups. Loran has been considered for this role, and, as this chapter shows, Loran's integration with the GNSS can yield other benefits.

This chapter gives an overview of why and how to integrate the GNSS with Loran. To this end, Sections 11.2 and 11.3 provide, respectively, an overview of Loran and a brief description of the theory behind Loran's operation. Then, Section 11.4 discusses the motivations for GNSS/Loran integration and describes potential system architectures for realizing this integration. Similar to the GNSS/INS integration architectures described in Chapter 6, the various GNSS/Loran integration methods have advantages and drawbacks. The remainder of the chapter describes the details of these integration schemes. This will include discussions of the advantages and disadvantages of the various integration architectures.

11.2 Loran Overview

Loran is a terrestrial, pulsed, LF, horizontal navigation system [1, 2]. Due to the nature of LF propagation and the power of the Loran transmission, its signals travel long distances. Users at distances of 800 km or more from a Loran transmitter can receive strong signals, thus enabling the long-range navigation.

11.2.1 Loran-C

Loran-C is the current operational form of Loran. There are 29 Loran-C stations in North America, providing coverage to CONUS, lower Canada, and a significant

portion of Alaska. Northern Europe currently has eight operating stations. Asia has numerous stations generally covering the Far East and Indian coastal area as well as Saudi Arabia. Russia operates a compatible system called Chayka. In all, Loran provides extensive coverage throughout the northern hemisphere. Figure 11.1 shows worldwide Loran coverage for both navigation (using a triad of stations) and precise time. More details on Loran can be found in numerous papers and books such as [1–3].

Loran-C has been an operational navigation system since January 1, 1958. It was initially developed primarily as a maritime navigation system, particular in the coastal confluence zone (CCZ). Traditional use of Loran relied on hyperbolic lines of positions (LOPs) derived from time differences (TD) of signal arrival from different stations. A TD from a station pair results in a hyperbolic LOP where the user can be potentially located. Horizontal positioning is accomplished by taking TD measurements from two different station pairs. This results in two LOPs with the user location being at the intersection of the lines. To assist positioning, maritime Loran charts provided the LOPs for various TD measurements. Figure 11.2 provides an illustrative example. Mariners would then derive TDs from their receivers (in the early generations, this involved reading measurements off a scope) and use the chart to determine the LOP corresponding to the measurement.

Over time, the Loran system and its usage have evolved. By the 1970s, spurred on by events such as the wreck of the *Argo Merchant*, Loran was mandated for carriage on tankers operating in U.S. waters. In the 1980s, Loran was evaluated for aviation navigation and approved as a supplemental navigation aid for en route navigation. The development of GPS changed the role and necessity of Loran. Overall, its use declined in favor of GPS. However, the advent of GPS also encouraged the development of complementary Loran functions.

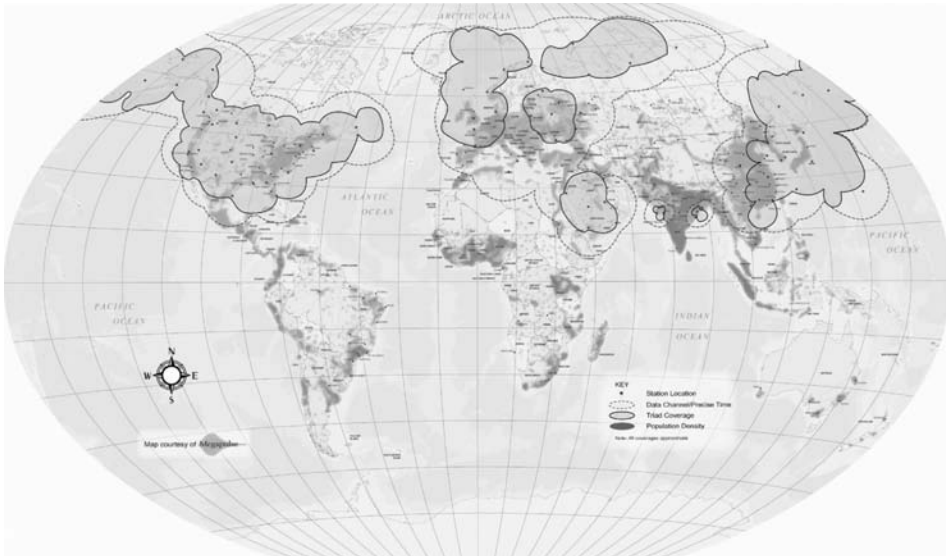


Figure 11.1 Loran worldwide coverage. (Courtesy of Megapulse. Reprinted with permission.)

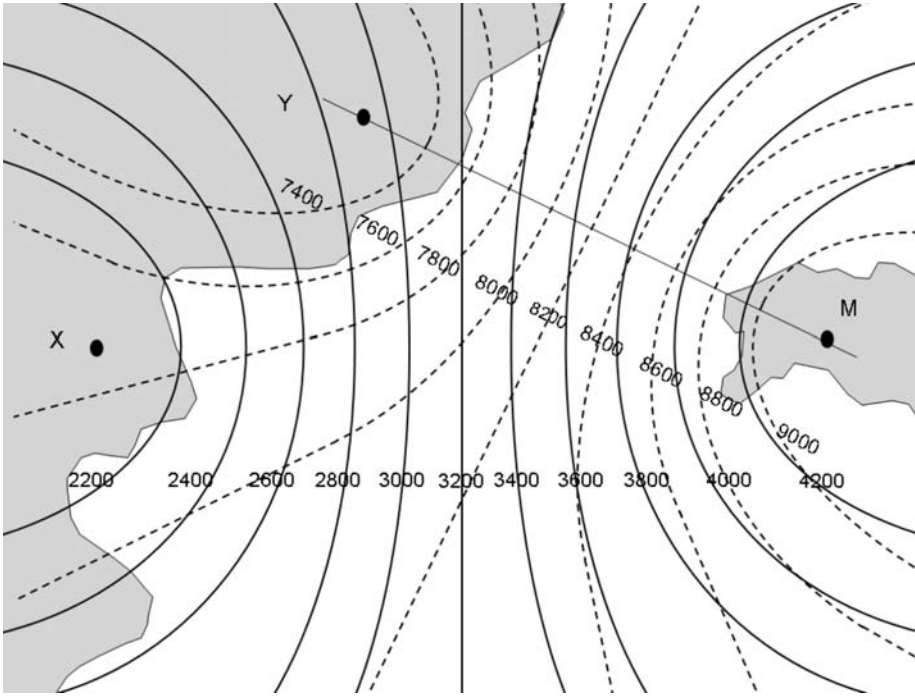


Figure 11.2 Example map with LOPs.

11.2.2 eLoran

eLoran is the next generation of the Loran system. It improves upon the Loran-C system through enhancements in equipment, transmitted signal, and operational concept. It is designed to support multiple modes of operation, including some of the most stringent position, navigation, and timing (PNT) applications [4]. eLoran can support aviation through the nonprecision approach (NPA) for landing at a required navigation performance of 0.3 (or RNP 0.3) and supports en route guidance at RNP 1.0. In addition, eLoran can support difficult maritime operations such as harbor entrance and approach (HEA) as well as maintaining CCZ operations. For timing and frequency, eLoran provides Stratum 1 frequency and highly synchronized (20-ns) time to Universal Time Coordinated (UTC). As such, the U.S. Department of Homeland Security (DHS) announced in February 2008 that eLoran would be implemented to provide “an independent national PNT system that complements GPS in the event of an outage or disruption in service [5].”

The most noticeable difference between the eLoran and Loran-C signal is the incorporation of a Loran data channel. This data channel provides the means of supporting the operations discussed above. It provides differential Loran corrections to support maritime, timing information for time and frequency users, and integrity warnings for aviation applications. In the United States, the data channel is provided on a modulated ninth pulse [6]. The design separates the modulated pulses from the navigation pulses. This is primarily done to aid aviation integrity.

eLoran is being developed as a global standard. The United States is working towards eLoran signal specifications, in coordination with Europe, which is also

developing eLoran. The two parties have spelled out their harmonized, high-level vision in the eLoran description document [7]. However, there may still be differences in the details. One difference that is likely to remain is the means of implementing data transmission on Loran. The European system will likely adopt Eurofix. The decision is based on both historical and operational reasons [21]. Europe is developing eLoran specifically for the maritime and timing applications. For these applications, the separation of data from navigation pulses is not as important a driver. Additionally, Eurofix has been in operation in Europe since 1997 and has roughly the same data capacity as ninth pulse.

With eLoran being designed specifically to provide operational backup to GNSSs, it is natural to expect navigators, especially in high-value applications, to carry both sensors. Additional benefits can be gained with intelligent integration of these sensors since the characteristics of the Loran signal complement those of GNSS. Since the characteristics of Loran differ from those of the INS, the integration techniques will differ from those used for INS/GNSS integration.

11.3 Theory of Operation

The Loran broadcast consists of a set of eight or nine shaped pulses on a 100-kHz carrier. Equation (11.1) gives the basic form of the pulse.

$$s(t) = pc \times t^2 \exp\left(\frac{-2t}{65 \mu\text{sec}}\right) \times \sin\left(\frac{2\pi t}{10 \mu\text{sec}}\right) \quad (11.1)$$

The signal may be transmitted in-phase or 180 degrees out-of-phase, equivalent to having the phase code (pc) set to 1 or -1 . Loran-C stations typically transmit a set of eight pulses at specified intervals with each pulse starting one millisecond after the start previous pulse. The pulsed nature of the system helps reduce propagation-related interference as explained later while the interval between pulse sets reduces intrasystem interference.

Loran stations are organized into regional groups of three or more stations forming a chain. Within a chain, each Loran-C station transmits its set of eight or nine pulses at the same regular interval known as the group repetition interval (GRI). The GRI of each chain is unique. There is one master station for each chain that, under Loran-C, sets the timing for the chain. The master station has an additional pulse that is transmitted 2 milliseconds after the start of the eight pulses to aid in identification. The other stations, known as secondary stations (or secondaries), transmit at predetermined intervals based on a published nominal emission delay (NED) relative to the master. This delay is created such that each signal has its own exclusive time window within the chain.

The master station broadcast differs in several ways from the secondary. First, in Loran-C, the master broadcasts nine pulses in each GRI, whereas each secondary broadcasts eight. Second, the master station can be uniquely identified based on the phase code sequence on the pulses, which differs from the phase code sequence on a secondary (whose phase code sequences are all the same). Note that the phase

code interval (PCI), the period over which the phase code sequence repeats, is two GRI.

Traditionally, position determination and navigation using Loran was accomplished using a single chain and measuring either TD or time of arrival (TOA). These two methods are also referred to as hyperbolic and pseudoranging methods, respectively. TD measures the difference in propagation time between two signals. Single-chain processing facilitates the measurement by enabling signal identification and calculation of relative propagation time. After the master signal has been identified, the identity of each secondary signal can be determined based on the time window discussed above. TD, in its simplest form, is then the difference in TOA between the master and a secondary after removing the NED of the secondary. Again, two independent TD measurements are thus adequate to determine horizontal position. TOA measurements are also straightforward when using a single chain. Pseudoranges akin to those of the GNSS can be determined from the TOA of the signal from each station by removing the station's respective NED. In the absence of propagation delays and transmission errors, each pseudorange differs from the true range by a common clock offset. This offset derives from several sources but is common to all measurements and can be solved for using the traditional iterative least-squares position estimate. When used in chain-based solutions, the TD and TOA solutions are essentially the same. Indeed, it can be shown that the TOA and TD methods are mathematically equivalent.

“All-in-view” operation is a term used to describe the ability to equally use any and all stations received regardless of which chain they are from. Position can be obtained with signals from multiple chains, though with additional requirements. One requirement is the ability to identify the transmitting station of the signals used from each chain. Under Loran-C, the identification of secondaries without other information, such the master station signal, is difficult. However, even with station identification, each chain can only be used individually, as in single-chain operations. In order to use signals across multiple chains, the relative time offset between the chains must be known. This can be addressed in multiple ways. It can be solved for as an additional unknown, particularly when using TOA or pseudoranges. It can be eliminated with the loss of one measurement. This is implicitly accomplished under TD when TDs are made only using measurements within a chain. These methods essentially require that each chain used contains at least two signals if there is to be a net benefit.

The above methods require signals from multiple stations in a chain for meaningful use. True all-in-view operation means being able to benefit from any available signal. This requires being able to identify the station and determine the chain offset. One method is to solve for the chain time offset by widelaning the signals from the different chains. The pattern of relative time differences between signals of two U.S. chains repeats roughly every 30–60s with the PCI being integer multiples of 200 microseconds. For Europe, chains repeat roughly every 300–600s. The European chains differ by a factor of 10 since the PCIs are integer multiples of 20 microseconds. If one has station identification and a reasonable estimate of position (~10 km for the United States, ~1 km for Europe), the chain time offset can be determined through widelaning. The implication of the position accuracy needed is that, in the United States, one can suffer a cycle slip and still accomplish

station identification. This is not possible in Europe. Having a good initial time estimate can be used to determine the offset without the need to widelane. Even a reasonable time estimate (within a few seconds) can narrow the scope of the widelane such that a position estimate is not necessary.

eLoran was designed to facilitate all in view operation. Hence it incorporates messages to identify each station and determine precise time. Additionally, it mandates time of transmission (TOT) control whereby each station is synchronized to UTC and broadcasts at a precisely known time. Under the current Loran-C system area monitor (SAM) control, the secondaries may transmit as much as 1 microsecond off the anticipated time based on the published NED. The result is that all-in-view operation with SAM control can suffer very large errors that do not exist under TOT control.

The design and transmission of the Loran signal is influenced by factors that affect LF signal propagation. LF signals can be received through two modes: ground wave and sky wave transmission. The ground wave signal can consist of both a surface and space wave. Generally, a terrestrial user receives the surface wave, which propagates along the Earth. As a result, the propagation speed and signal attenuation is affected by Earth properties such as ground conductivity and terrain roughness. Some of these properties are static in time while others vary temporally. This leads to propagation delays and variations from that nominal delay. Numerous models have been created to estimate the delay and attenuation [8, 9]. The sky wave occurs when the signal propagates up through the atmosphere and reflects or “hops” off the ionosphere back to the user. The sky wave signal is more inconsistent as its signal strength and propagation time varies significantly. The variation is more pronounced than the ground wave propagation variation due to the daily and hourly fluctuations of the electrical characteristics of the ionosphere. Indeed, for navigation, Loran sky wave is considered a nuisance as it interferes with the more stable ground wave signal. Complicating matters is that the sky wave can be stronger than the ground wave and received at longer ranges. Several aspects of the Loran signal design (fast rise time of the Loran envelope, phase coding) were incorporated as mitigation strategies for sky wave inference.

In Loran, the ground wave delays are divided into three major propagation delay factors. The primary factor (PF) accounts for the propagation time needed to traverse the atmosphere. The secondary factor (SF) is the increment of time for traversing an all seawater path. Both the PF and SF are calculated based on standard models and are fixed for a given signal at a given location. Additional SF (ASF) encompasses the remaining delay—that is, the extra delay on the Loran signal due to propagation over nonhomogenous land path instead of an all seawater path. ASF represents the largest source of variation in the Loran measurements. As a result, ASF estimates are traditionally used in Loran, and quoted accuracies usually assume use of ASF maps, with static estimates for ASF. However, even if such maps are used, the residual ASF (i.e., its temporal variation from the static ASF estimate) can be significant and as large as 500m or more peak to peak. The seasonal variations are greatest when there are changes in temperature and moisture content that are significant as seen in places such as the northeast of the United States. These changes affect ground conductivity. Hence, weather has a strong effect on variations in ASF [10]. Large spatial changes can occur in areas of

rough terrain such as mountainous regions [8, 11]. Also, ASF varies spatially with elevation; namely it decreases with increasing height above the topography. To enable maritime HEA, differential Loran corrections broadcasts specifically treat this temporal variation allowing for sub-10-m accuracy in position. For aviation, such corrections are not employed and so a model bounding these ASF temporal variations is used. Providing corrections with the necessary integrity and coverage for aviation was thought to be too cost-prohibitive for a backup system.

11.4 Historical Reasons for GNSS/Loran Integration

Ideas on integrating Loran and GNSS have been discussed since the 1980s [12]. The primary motivator for using the combination was due to the limitations of the GPS constellation. The limited number of visible GPS satellites led to the idea of combining GPS and Loran pseudoranges. Details of how this is accomplished are presented later in Section 11.5.2 and can be found in [6]. Another reason for combining Loran and GPS stemmed from SA. The idea was to use the better short-term repeatable accuracy of Loran with the inherently better long-term accuracy of GPS [13].

By the 1990s, improvements in the GPS constellation and the elimination of SA reduced interest in integrated Loran/GPS. Only in certain applications where GPS signals are limited or weak was such a combination considered of practical utility. An example of such an application is navigation in urban canyons where this environment reduces the visibility of GPS satellites. Additional pseudoranges from Loran can improve navigation performance and availability in this application [14]. Another example is leveraging the accuracy of Loran timing to aid reception of GPS signals indoors.

A more recent motivation for this integration is the need for redundancy, particularly for safety-of-life applications. Integrating Loran and GNSS for safety and redundancy is attractive because they have independent failure modes and complementary features. While the mixing of information is inherent in integration, this must be done with consideration. A guiding principle in safety-of-life integrity is the separation of hazards so that they can be properly accounted for. Integration, if not done carefully, can result in the hazards of one system affecting the other. So the challenge is to combine information in such a way that faults either do not “cross over” or are bounded.

One limitation on the use of Loran is that it is a horizontal positioning system. Loran ranges are measurements of the distance between the user and transmitter over the surface of the Earth [21]. It is not known if the range has any dependency on altitude when near the Earth’s surface (<5 km). Fortunately, this range difference is on the order of tens of meters, at most. However, this can be significant for aviation integrity and some care should be taken when combining it with GPS for vertical position. Hence the most natural combination for safety of life of GPS/Loran is for land or maritime applications that occur on the Earth’s surface.

11.5 Integration Scenarios

The integration techniques used for Loran/GNSS can be categorized in several ways. For the purpose of our discussion, we have chosen to classify the integrations into three categories based on the “depth” of integration: (1) position domain, (2) range domain, and (3) tracking domain. Of course, there are some integration techniques that span two domains. These instances are discussed in the context of the deeper or “most fundamental” form of information used. For example, the use GNSS position with Loran range measurements is considered as an example of range domain integration. The level of integration achievable is dependent on the equipment available. This potential ambiguity in classifying integration schemes is similar to the information presented in Chapter 6 on GNSS/INS integration.

In examining each integration technique, the focus is on desired capabilities. Common goals are improving availability, accuracy, or integrity. So, for example, range domain integration is examined from the perspective of improving accuracy or integrity. Related to the integrity goal is the mitigation of threats to the primary navigation aid, which in this case is the GNSS. These threats are related to different magnitudes of GNSS signal, which can range from complete loss of signal (jamming), to limited signal availability (obscuration), or weak signals, as described in Chapter 12.

11.5.1 Position-Domain Integration

Position-domain combination provides an easy means of integrating GNSS and Loran, even if these systems are two separate pieces of equipment. This form of integration is seen in Figure 11.3. It allows both systems to operate independently and requires only the most basic sensor outputs. In fact, each system can be a separate and independent unit as long as each unit has an output that can be referenced to a common time standard at a reasonable output rate. The integration does not place substantial requirements on the internal operations of each receiver. For example, Loran positioning can be derived from either TD or TOA (Loran-range) measurements. The drawback is that the benefits of this form of integration are minimal when both systems are operating nominally. This is because GNSS position domain performance is typically much better than that of Loran. The most

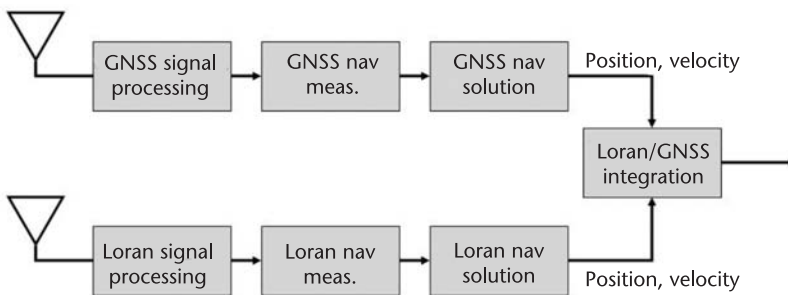


Figure 11.3 Position-domain integration.

important benefit of this integration is having two PNT sensors that can operate independently.

An easy way of integrating GNSS and Loran in the position domain is using a snapshot solution. This can be implemented using traditional techniques such as minimum variance estimation. Equations (11.2)–(11.4) are the basic minimum variance combination equations.

$$w_{Loran} \mathbf{p}_{Loran} + w_{GPS} \mathbf{p}_{GPS} = \mathbf{p}_{tot} \quad (11.2)$$

$$w_{Loran} = \frac{\sigma_{GPS}^2}{(\sigma_{Loran}^2 + \sigma_{GPS}^2)} \quad (11.3)$$

$$w_{GPS} = \frac{\sigma_{Loran}^2}{(\sigma_{Loran}^2 + \sigma_{AGPS}^2)} \quad (11.4)$$

The addition of Loran in a minimum variance context is limited for two reasons. First, the Loran position input provides very little value added since its uncertainty is generally far larger than that of GNSS. That is, σ_{Loran}^2 is much larger than σ_{GPS}^2 . As seen in the equation, this translates to a total or blended solution that heavily weighs the GNSS output. Second, the error on stand-alone Loran violates the assumption of random (unbiased) error that is the basis of minimum variance. As a result, the addition of Loran, without reasonable ASF measurements, may not actually help the solution [22].

Filtering approaches such as those used with GNSS/INS integration discussed in Chapter 6 can mitigate the effects of ASF in the position-domain integration scheme. This is because they consider a series of measurements from several epochs in generating a position solution. Since ASFs do not change quickly in time, their effect in the position domain will not vary much provided that the same stations are used in the solution. The low temporal variation of ASF results in Loran having high repeatability. This characteristic of Loran can be utilized for position-domain integration. A simple method is to integrate based on the difference of the solutions from two consecutive measurements. This differencing eliminates the common-mode ASF and can be used to extrapolate from the previous position solution. Another method is to use an estimation filter to estimate the position bias in the Loran solution, which is dominated by the ASF effects. The problem with any of these techniques in the position domain is that the bias estimates have relevance only near the location they were estimated and with the same combination of Loran stations used to generate the estimates. Estimating ASF in the range domain rather than its position domain provides a more useful solution.

The integrity benefit achievable through position-domain combination is minimal. Comparing position solutions from Loran and GNSS does not provide an easy means of detecting fault because (1) the Loran solution is generally less accurate, and (2) if the two solutions disagree, one cannot decide which solution, if any, to trust. If the GNSS fault is large and systematic (for example, spoofing), having an extra independent system such as Loran may be of some utility for

detection. Loran has greater spoofing resistance due to features of the system and its signal [22].

In summary, the benefits of position-domain integration reside mostly in increased availability and the redundancy of having two independent systems. Accuracy and integrity benefits are limited due to the disparity in the performance of GNSSs and Loran.

11.5.2 Range-Domain Integration

Range-domain integration of GNSS and Loran provides much more powerful uses of the combination with a level of hardware integration and complexity comparable to position domain. This form of integration is illustrated in Figure 11.4. Range-domain combination provides the potential for ASF calibration and better cross-checking of measurements. Thus, it can be used to improve the availability, accuracy, and integrity of the combined solution. Use of the range domain has higher requirements than the position domain since TOA measurements are necessary. TOAs are necessary to derive the Loran range measurements. For the purposes of this discussion, the term TOA will be used both to refer to the time estimate and the estimated Loran range derived from TOA. TOT control is also desirable. As previously mentioned, without TOT control, we have to solve for a clock offset. Under TOT, only one Loran clock offset needs to be solved. In what follows we consider a basic range-domain integration scheme. The most basic range-domain integration takes the range measurements from GNSS and Loran and combines them into one set of navigation equations. The basic equation used is (11.5), which is a linear equation similar in form to (3.7) discussed in Chapter 3.

$$\mathbf{z} = \begin{bmatrix} \delta\rho_1 \\ \vdots \\ \delta\rho_N \\ \delta d_{TOA,1} \\ \vdots \\ \delta d_{TOA,M} \end{bmatrix} = \mathbf{G} \begin{bmatrix} \delta\mathbf{x} \\ \delta b \\ \delta d \end{bmatrix} + \epsilon \quad (11.5)$$

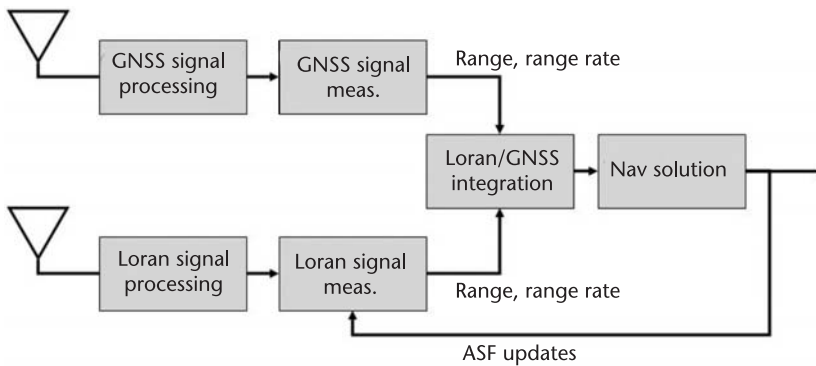


Figure 11.4 Range-domain integration.

The vector $\delta \mathbf{z}$ left-hand side of the equation is constructed using GNSS pseudo-range measurements that are N in number and Loran TOA measurements that are M in number. More specifically,

$$\delta \mathbf{z} = \begin{bmatrix} \delta \rho_1 \\ \vdots \\ \delta \rho_N \\ \delta d_{TOA,1} \\ \vdots \\ \delta d_{TOA,M} \end{bmatrix} = \begin{bmatrix} \rho_1 - \hat{\rho}_1 \\ \vdots \\ \rho_N - \hat{\rho}_N \\ d_{TOA,1} - \hat{d}_{TOA,1} \\ \vdots \\ d_{TOA,M} - \hat{d}_{TOA,M} \end{bmatrix} \quad (11.6)$$

The variables with “hats” above them are estimated quantities computed based on our current best estimation of position. The matrix \mathbf{G} is the geometry matrix but has a slightly different form from the one given in Chapter 3. More specifically, the matrix has the following form:

$$\mathbf{G} = \begin{bmatrix} G_{1,1} & G_{1,2} & G_{1,3} & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ G_{N,1} & G_{N,2} & G_{N,3} & 1 & 0 \\ G_{N+1,1} & G_{N+1,2} & G_{N+1,3} & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ G_{N+M,1} & G_{N+M,2} & G_{N+M,3} & 1 & 1 \end{bmatrix} \quad (11.7)$$

The upper-half of the matrix is identical to the geometry matrix given in Chapter 3, while the lower-half is the geometry matrix for the Loran problem. The unknown states that need to be solved for consist of position difference vector ($\delta \mathbf{x}$) and the receiver clock bias relative to GNSS time (δb) as discussed in Chapter 3. In addition, however, we must solve for the difference between the GNSS and Loran time (δd). Note that time δd accounts for both differences in GNSS and Loran system time as well as offsets between the GNSS and Loran receiver clocks in the case that a common clock is not used. A different formulation is also possible with δb and δd representing the GNSS and Loran time offsets from absolute time, respectively. In this case, the fourth column of the geometry matrix \mathbf{G} is zero for rows $N + 1$ to $N + M$ (i.e., rows associated with the Loran measurements). Note that this implicitly assumes that only one GNSS system is used. If more than one GNSS is used, then the system time differences between the GNSSs must also be accounted for.

Before solving (11.5) we need to apply one of several corrections to the measurement. These include differential corrections for GNSS as well as PF, SF, and ASF terms for Loran. Since these corrections are typically location dependent, they are applied as one iterates to solve the navigation equations. Non-location-dependent corrections (such as ionosphere from dual-frequency measurements) should be applied prior to forming the corrected pseudoranges or TOAs. Additionally, when generating the Loran range estimate \hat{d}_{TOA} , distance over the surface of the Earth instead of the Cartesian distance should be used.

Equation (11.5) is solved using an iterative, weighted, least-squares approach as given by (11.8):

$$\begin{bmatrix} \delta \mathbf{x} \\ \delta b \\ \delta d \end{bmatrix} = \mathbf{S} \delta \mathbf{z} \quad (11.8)$$

where the matrix \mathbf{S} is given by:

$$\mathbf{S} = (\mathbf{G}^T \mathbf{W} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{W} \quad (11.9)$$

where \mathbf{W} is the weight matrix. Equal weighting is generally not recommended because the GNSS measurements tend to be significantly better than the Loran measurements. A common weighting used is based on error-covariance matrices R for the GNSS and Loran measurements as given below:

$$\mathbf{W}^{-1} = \begin{bmatrix} R_{GNSS} & 0 \\ 0 & R_{Loran} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \sigma_N^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{N+1}^2 & \dots & \sigma_{N+1, N+M} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \sigma_{N+1, N+M} & \dots & \sigma_{N+M}^2 \end{bmatrix} \quad (11.10)$$

While using a covariance-based weighting provides better performance, the result is only optimal if good covariance estimates are available and there is little bias. Typically, GNSS errors are considered uncorrelated and hence the off-diagonal terms related to GNSS can be approximated by zero. These errors generally have little to no bias component. However, to use the Loran measurements in a reasonable way, good ASF estimates are needed. Without good ASF estimates, the Loran TOA will contain a potentially significant ASF. The significance of this is that ASF is a bias and has correlated components between the measurements from different stations at a given epoch. The weighted least-squares solution is not optimal for treating biased errors. With good ASF estimates, the cross-correlation and biases will be minimized.

This form of integration can clearly provide increased availability. Even if each system has three or fewer signals only and hence no three-dimensional position solution, the combination of signals from both systems can result in an available solution. Furthermore, the described integration only considers combining the measurements at one point in time. We can extend the integration over multiple periods through Kalman-filtering. In this case, we can benefit from the good repeatable accuracy of Loran to aid the overall solution; this can aid accuracy.

The range-domain integration can aid accuracy in several ways. However, improving accuracy through integration generally depends on having good ASF estimates or reducing the need for ASF (i.e., through the use of Loran measurements over multiple periods). The latter can be accomplished using an estimator or filter. The former can be achieved in several ways. One means is to have ASF maps and differential Loran (D-Loran) corrections, as envisioned in the use of E-Loran for HEA. Another means is to use the integration to provide ASF estimates for later use. An example of this method is to have an integrated Loran/GNSS dynamically generate Loran ASF maps and time offsets. This provides a similar function to D-Loran corrections [15]. This dynamic generation can be used when traversing the same area in a later journey—hence we refer to it as *déjà vu* navigation. Section 11.5.3 details the steps for this dynamic generation for illustrating Loran/GNSS range-domain integration. There are other ways to use integrated Loran/GNSS to estimate ASF. Estimation filters can estimate the ASF or it can estimate parameters for ASF models. A form of the latter and empirical models for ASF are presented in [16]. Having accurate ASF estimates improves Loran stand-alone accuracy and allows it to provide benefits, such as improved accuracy and integrity, to an integrated Loran/GNSS system. These integrity benefits are discussed in Section 11.5.4.

11.5.3 Déjà Vu Navigation: A Case Study of Range-Domain Integration

As mentioned previously, calculating the ASF grid can be achieved in many ways. The *déjà vu* navigation technique generates a full grid based solely on measurements taken in the course of normal operations. As more measurements and trips over the region are taken, the grid is refined and updated to improve its performance.

The first step is to collect calculated ASF and generate an updating ASF grid. The basic concept is seen in Figure 11.5. First measurements are collected from a Loran and GNSS receivers. If integrity is required, both the Loran measurements

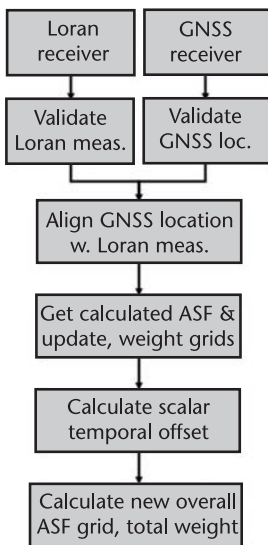


Figure 11.5 Generation of each calculated ASF.

and GNSS location should be validated. The Loran measurement should be free of cycle slips and nonnominal errors. The GNSS position needs integrity as it forms the basis for generating the ASF estimate. Integrity can be obtained via the use of GNSS integrity signals such as GBAS and SBAS, respectively. Next, the outputs of the GNSS and Loran are aligned such that they reference the same time and location. ASF estimates can then be calculated using GPS position and Loran range measurements. The GNSS position provides the true range, which is then converted to an “ASF-free” Loran propagation time by accounting for the PF and SF. This ASF-free propagation time is differenced from the measured propagation time to get the calculated ASF (in units of time). This is done for each Loran signal used. Mathematically, this is given by (11.11):

$$\begin{aligned} \text{Calculated ASF} = & \text{measured propagation time} \\ & - \text{“ASF-free” propagation time} \end{aligned} \quad (11.11)$$

Each calculated ASF is used to generate the entire grid—an updating ASF grid. This is achieved by having each calculated ASF contribute to the estimate of ASF at each grid point. The contribution or weight should, of course, depend on the relevance of the ASF estimate. In the technique tested in [15], an exponential weighting based on distance was used. This is seen in (11.12) where w_i represents the weight on grid point i , d_i is the distance from the measurement point to the grid point, and k is a constant.

$$w_i = k \exp\left(\frac{-d_i}{\text{gridsize}}\right) \quad (11.12)$$

A series of calculated ASFs can be used to generate a combined weighted ASF estimate by combining the updating ASF grid and weighting associated with each of them. Calculation of the temporal offset, especially if the measurements are taken apart in time, may be necessary for the combination. A second necessary component for having an accurate ASF is to have a term accounting for the temporal variation. This is necessary under nominal conditions for updating the grid. The ASF may have changed significantly between trips. It is also necessary to use the grid should GNSS be lost.

Under nominal conditions, the calculation of temporal terms can be achieved in many ways. One method is to compare the calculated ASF at the current point with the estimated ASF from the overall ASF map at the closest grid points. A more sophisticated technique also can be used. The calculated ASF, as discussed previously, is used to generate an ASF grid, represented in (11.13) in matrix form as $A_{ASF, curr}$.

$$\begin{aligned} \text{offset}_{temporal} = & \sum_{\text{term by term}} (A_{ASF, curr} - A_{ASF, map}) \\ & \times \frac{(W_{ASF, curr} \cdot W_{ASF, map})}{|(W_{ASF, curr} \cdot W_{ASF, map})|} \end{aligned} \quad (11.13)$$

Associated with the grid is a weighting matrix, $W_{ASF, curr}$, with weights corresponding to each point. The base ASF grid and weight matrix are denoted by $A_{ASF, map}$ and $W_{ASF, map}$, respectively. The temporal offset is then derived from the difference of two ASF maps multiplied entry-wise (Hadamard product) by the normalized weighting matrices. The process above shows how to get the temporal offset given by GNSS. However, the benefit of integration through the use of the ASF estimates requires that the temporal offset can be calculated during a GNSS outage. Several methods have been proposed for this. If the GNSS outage occurs in the course of navigation, then the last calculated offset is used. If the outage occurs prior to travel, a known location (i.e., a dock or airport hangar) can be used as a position reference instead of GNSS to generate the offset. The approach of using known position references can be applied in another way. If the travel should transit over known locations such as an area supported by eLoran HEA, this can be used for determining the offset. However, there are some caveats to the approaches. First, the offset will obviously degrade and decorrelate temporally. Second, the location used to generate the offset must be part of the ASF grid previously generated.

The final step is to incorporate the updating grid into the full grid map to generate a new grid map and weighting. The combination used is based on weighting. The calculation for the overall ASF map and weighting are seen in (11.14) and (11.15). The ASF estimate is generated by interpolating from the overall ASF map using the weights.

$$A_{ASF, map}^{new} = \frac{(A_{ASF, curr} \cdot W_{ASF, curr} + A_{ASF, map} \cdot W_{ASF, map})}{W_{ASF, map}^{new}} \quad (11.14)$$

$$W_{ASF, map}^{new} = W_{ASF, curr} + W_{ASF, map} \quad (11.15)$$

GPS and Loran data was collected and assessed to illustrate the basic steps of the algorithm. This is shown in Figure 11.6. The top left plot of the figure shows an updating ASF grid ($A_{ASF, curr}$), while the top right one shows its associated weighting ($W_{ASF, curr}$) generated during one trip. This grid and the weightings used to update the overall ASF grid ($A_{ASF, map}$) are shown in the bottom left map while the weighting ($W_{ASF, map}$) is shown in the bottom right. The overall maps were generated using roughly 30 trips from 2006 to 2008 in New London, Connecticut.

11.5.4 Integrity with Range-Domain Integration

Providing integrity requires, in part, determining if the signals used are performing within specified tolerances. This can be checked by using an overdetermined solution that combines the range measurements from both systems. Receiver Autonomous Integrity Monitoring (RAIM), for example, can be used for such purpose. One method starts with the integration equations (11.5)–(11.8). In this RAIM technique, one can examine the residuals of the final solution. The residuals provide a means of hypothesis testing between the no-fault case and faulted conditions. For identifying the error, one can examine the residuals of position solutions from different subsets of range measurements and perform similar hypothesis testing.

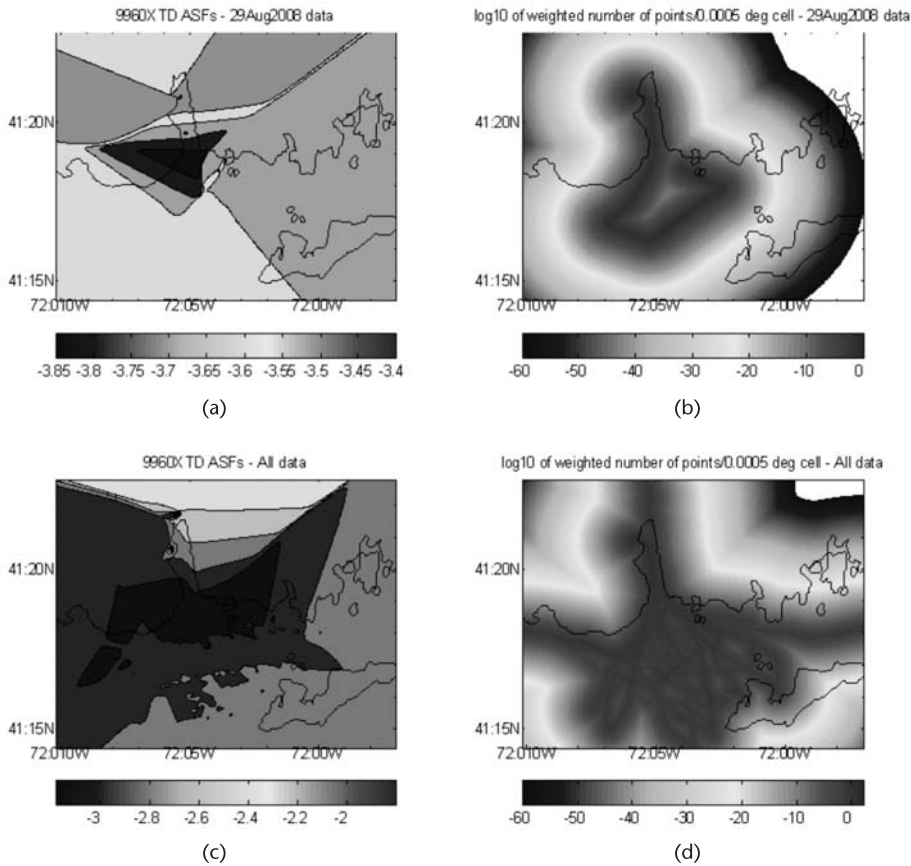


Figure 11.6 Updating ASF grid and weight (a, b) and overall ASF grid and weight (c, d).

Another technique is to use multiple hypothesis solution separation (MHSS) [17, 18]. Figure 11.7 illustrates the idea of hypothesis testing. The test statistics have different distributions in the no fault and faulted cases. A threshold is set such that if the test statistic is above the threshold, then the results are considered faulted; below the threshold, it is considered not faulted.

However, the fault detection and exclusion (FDE) is limited by the quality and characteristics of the input measurements. The nominal error on a Loran signal, particularly a weak (far-off) signal, can exceed the error of a faulted GNSS measurement. Additional concerns are with the bias and cross-correlation both primarily due to ASF in the Loran signal. The bias is generally larger for far-off signals. This is problematic as a common assumption in many RAIM algorithms is that there is little or no bias in the no-fault measurement with the biases coming in as a result of a fault. While techniques can be modified to treat nominal conditions with bias (such as [19]), the inclusion of biases in nominal measurements will decrease our ability to detect GNSS faults. As a result, we may want to limit the Loran signals used to only the closest (highest quality) signals when using them to aid in fault detection for GNSS. On the other hand, GNSS measurements can be useful in determining Loran faults or cycle errors in tracking Loran.

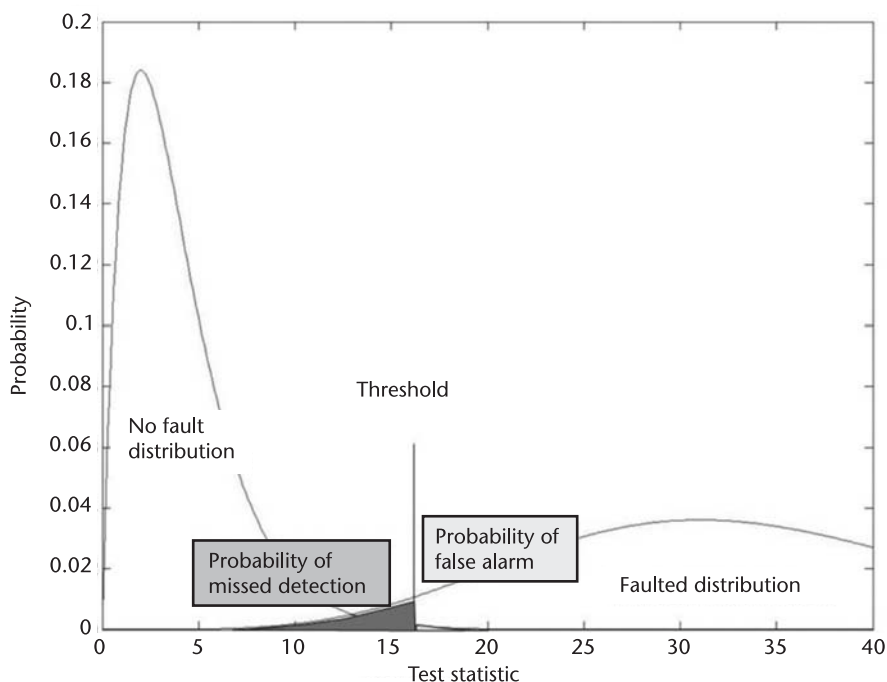


Figure 11.7 Basic hypothesis testing.

The solution is to have accurate ASF estimates that allow more Loran measurements to be beneficial for RAIM. A preliminary study of the benefits of the combination used in a MHSS RAIM algorithm for providing aviation integrity is given in [15]. The results indicate that being able to bound one standard deviation error on Loran to 10m has clear benefits while having a larger bound such as 40m has little benefit. The former is a case where there are good ASF estimates, and the later is one where there are only rough ASF estimates.

Integrated Loran/GNSS offers multiple options for providing integrity. A robust architecture is illustrated in the flowchart in Figure 11.8 where different fallbacks are presented. Nominal operations would use GNSSs with integrity (via an augmentation system or RAIM) to provide position and to generate ASF estimates (i.e., grids). Should stand-alone GNSS integrity be unavailable, then the combination

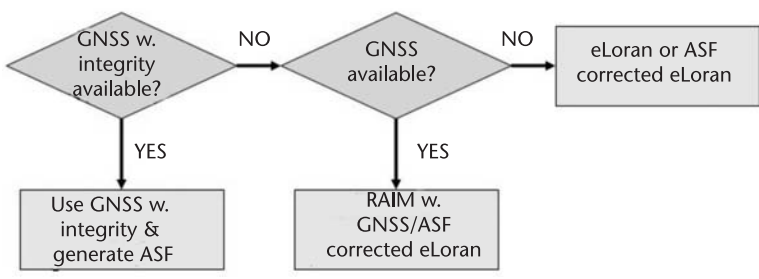


Figure 11.8 Integrity options for an integrated system.

of ASF-corrected Loran and GNSSs can be used to provide integrity through a combined system RAIM. Should GNSS be completely unavailable, E-Loran with ASF corrections can be used, perhaps with Loran integrity provided via a RAIM algorithm.

However, GNSS/Loran integration for safety-of-life applications must be embarked upon with care. The ASF estimate itself must have integrity. The residual errors on the estimate must be bounded. For example, if GNSS is used to determine the ASF, the position error on GNSS must be bounded. Techniques like déjà vu navigation used the integration to generate the ASF estimates. Hence, to protect integrity and prevent the feed-through of faults, GNSS integrity must first be assured before it is used to generate ASF. Indeed, a guiding principle for these applications is the separation of operations so that faults from one system do not affect the other systems. Section 11.5.5 discusses how to get accurate ASF with integrity.

11.5.5 Improved Accuracy for Loran Integrity

ASF estimates generated from GNSSs can be used provided two caveats are followed. First, the GNSS measurement used to estimate ASF must have integrity. This means that a GNSS integrity signal (i.e., SBAS or GBAS) must be used to generate the solution. For the GNSS signals to truly have integrity, the GNSS integrity system must not have alarmed within the TOA of the integrity system. For example, that means that we cannot use the ASF generated from a given GNSS measurement unless the integrity signal has not alarmed for a length of time equal to the time to alarm (TTA) after the measurement. Second, an integrity model for the degradation of ASF estimates in time and space is required. ASF spatial and temporal models can be used to grow the error bound on the ASF [10, 11]. A basic model, given in (11.16) can be expressed as having the position domain ASF bound be the sum of the nominal bound and factors ($k_{ASF,d}d$, $k_{ASF,t}t$) that account for the growth of ASF in distance and time. $k_{ASF,d}$ and $k_{ASF,t}$ are bounding growth factors in distance and time, respectively.

$$ASF_{bound} = ASF_{bound,nom} + k_{ASF,d}d + k_{ASF,t}t \quad (11.16)$$

The most important and difficult step is validating the bounding growth factors. A related area to navigation integrity is signal integrity. That is, can we develop the ability to detect spoofing and other attacks on the signal? The integration of GNSS and Loran can be used to enable antispoofing and authentication. Loran has many features that aid this process. First, it is difficult to spoof Loran. The equipment required to transmit a Loran signal is large and requires a significant amount of power. Even if an attacker could build and operate a spoofing transmitter without detection, a Loran H-field antenna could determine the relative directions of incoming signals. Consistency checks on direction and relative signal power would force an attacker to use a multiplicity of transmitters to even attempt spoofing. Finally, it is possible that eLoran will incorporate authentication messages making spoofing Loran even more difficult [20]. This confidence that Loran signals cannot be spoofed can be leveraged to authenticate GNSS signals.

11.5.6 Tracking Loop Domain Integration

Integration of Loran and GNSS can be useful at the more fundamental level of signal acquisition and tracking. An example of the architecture is seen in Figure 11.9. Loran measurements can be received in environments such as deep urban canyons, parking structures, and indoor locations, where GNSS signals are very weak. These measurements from Loran can be used in aiding the acquisition and tracking of the GNSS signals whereas stand-alone GNSS would not be able to operate.

High-accuracy time and frequency measurements from Loran provide a means to enable longer integration periods for GNSS signal acquisition. Extending the integration period enhances the SNR, where the maximum period for noncoherent integration, T , is a function of the stability of the clock used by the GNSS receiver and the chipping rate (chip length) of the desired signal. The target is to keep the deviation within half a chip. This is shown in (11.17) where c = chip time [1 microsec for C/A, 100 ns for P(Y), L5, etc.] and s is the clock stability in fractional form (1 ppm = 1/1000000).

$$T \leq \frac{\left(\frac{1}{2}c\right)}{s} \quad (11.17)$$

So given 1-ppm clock stability, this implies $T = 0.5$ second for a 1-MHz chipping rate and 0.05 seconds for a 10-MHz chipping rate. This is a typical order of magnitude for a crystal oscillator. Given the clock stability using Loran should be on the order of 10 ppb, one can realize a significant increase in integration time and hence improvement of SNR. The enhancement is significant particularly for users of 10-MHz codes such as the P(Y) or L5 code. Hence such an integration may be particularly useful for military receivers operating in urban and indoor environments.

Since power improves as the square root of the increase in integration time, the potential SNR enhancement is given by (11.18):

$$\Delta \text{SNR} = 10 \log_{10} \left(\frac{T}{t_{\text{nom}}} \right)^{1/2} = 5 \log_{10} \left(\frac{\frac{1}{2}c}{s \times t_{\text{nom}}} \right) \quad (11.18)$$

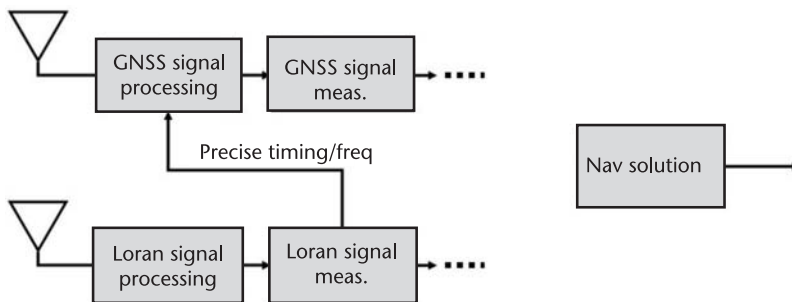


Figure 11.9 Tracking loop integration.

11.6 Conclusions

This chapter discusses the various available architectures for integrating GNSSs with Loran. The complementary nature of the two systems allows for an integrated solution that can improve navigation availability, robustness, or even integrity. Position-domain integration architectures are reasonably easy to implement and can provide some benefits. However, these benefits are marginal, especially compared to the gains possible in range-domain integration. For many applications, integration using range-domain measurements offers the desired performance without significantly more complexity. It offers reasonable ways of estimating Loran ASF and creating a combined system that has significant performance benefits over either stand-alone system. Deeper integration at the tracking loop level may be used to improve GNSS acquisition by enabling longer GNSS integration times.

References

- [1] Forssell, B., *Radionavigation Systems*, Englewood Cliffs, NJ: Prentice Hall, 1991.
- [2] *Loran-C User Handbook*, COMDTPUB P16562.6, November 1992.
- [3] Frank, R. L., "Current Developments in Loran-C," *Proceedings of the IEEE*, Vol. 71, No. 10, October 1983, pp. 1127–1139.
- [4] FAA report to FAA Vice President for Technical Operations Navigation Services Directorate, "Loran's Capability to Mitigate the Impact of a GPS Outage on GPS Position, Navigation, and Time Applications," March 2004.
- [5] Press Office, U.S. Department of Homeland Security, "Statement from DHS Press Secretary Laura Keehner on the Adoption of National Backup System to GPS," February 7, 2008.
- [6] Lo, S., et al, "Loran Data Modulation: Extensions and Examples," *IEEE Transactions on Aerospace and Electronic Systems*, Volume 43, Issue 2, April 2007 pp. 628–644.
- [7] International Loran Association, "Enhanced Loran (eLoran) Definition Document," version 0.1, January 2007.
- [8] Samaddar, S. N., "The Theory of Loran-C Ground Wave Propagation—A Review," *Navigation: The Journal of The Institute of Navigation*, Vol. 26, No. 3, Fall 1979.
- [9] Johler, J. R., W. J. Kellar, and L. C. Walters, "Phase of the Low Radiofrequency Groundwave," Circular 573, National Bureau of Standards, United States, 1956.
- [10] Samaddar, S. N., "Weather Effect on Loran-C Propagation. Navigation," *Journal of the Institute of Navigation*, Vol. 27, No. 1, Spring 1980, pp. 39–53.
- [11] Peterson, B., et al, "Analysis of Terrain Effects on DGPS and LORAN Signals," *Proceedings of the Institute of Navigation National Technical Meeting*, Anaheim, CA, January 2000.
- [12] Peterson, B., et al, "Integrated GPS/Loran: Structures and Issues," *Navigation: The Journal of the Institute of Navigation*, Vol. 45 No. 3, 1998.
- [13] Enge, P. K., and J. R. McCullough, "Aiding GPS with Calibrated Loran-C," *Navigation: The Journal of the Institute of Navigation*, Vol. 35 No. 4, 1988.
- [14] Carroll, J., et al, "Use of an Integrated GPS/Loran Tracking System in the Urban Environment," *Proceedings of the Institute of Navigation Annual Meeting*, Cambridge, MA, June 2005.
- [15] Peterson, B., S. Lo, and P. Enge, "Integrating GPS and Loran for Integrity in Safety of Life Applications," *Proceedings of the Institute of Navigation GNSS Conference*, Savannah, GA, September 2008.
- [16] Pisano, J. J., et al, "Using GPS to Calibrate Loran-C," *Proceedings of the Institute of Navigation GPS Conference*, Colorado Springs, CO, Sept. 1990.

- [17] Pervan, B. S., S. P. Pullen, and J. R. Christie, "A Multiple Hypothesis Approach to Satellite Navigation Integrity," *Navigation: The Journal of the Institute of Navigation*, Vol. 45 No. 1, 1998.
- [18] Blanch, J., et al, "An Optimized Multiple Hypothesis RAIM Algorithm for Vertical Guidance," *Proceedings of the Institute of Navigation GNSS Conference*, Fort Worth, TX, September 2007.
- [19] Lo, S., B. Peterson, and P. Enge, "Proving the Integrity of the Weighted Sum Squared Error (WSSE) Loran Cycle Confidence Algorithm," *Navigation: The Journal of the Institute of Navigation*, Vol. 54 No. 4, 2007.
- [20] Qiu, D., et al, "Geoencryption Using Loran," *Proceedings of the Institute of Navigation National Technical Meeting*, San Diego, CA, January 2007.
- [21] van Willigen, D., "Eurofix," *Journal of the Royal Institute of Navigation*, Vol. 42, No. 3, September 1989.
- [22] Lo, S., Peterson, B., and P. Enge, "Assessing the Security of a Navigation System: A Case Study Using Enhanced Loran," *European Navigation Conference GNSS 2009*, Naples, Italy, May 2009.

Indoor and Weak Signal Navigation

Heidi Kuusniemi and Timo Jokitalo

12.1 Introduction

In recent times, the GPS receiver market has grown significantly, and the amount of services has increased greatly, partly due to the legislative cellular E-911 and E-112 demands (e.g., [1, 2]). This together with the consumer demands for better positioning availability drive the need to extend the positioning capability of receivers into ever weaker signal environments. Commercial location-based services are most often accessed in notoriously difficult signal environments, such as urban canyons as well as indoors. Thus, there is a growing need to track continually lower power signals and to produce a robust user PVT solution even when the measurements are very noisy. This requires various issues to be taken into account when stand-alone GNSS positioning is considered—both in signal acquisition and tracking as well as in navigation algorithms and quality monitoring. The elements that have to be taken into consideration during indoor and weak signal GNSS positioning form this chapter. The chapter also covers assistance information, like time and satellite orbit parameters, as well as supporting systems for navigation (e.g., pseudolites, self-contained sensors, and WLAN), which benefits indoor and weak signal positioning. The aim of this chapter is to provide an informal overview that is accessible to those lacking a theoretical background in the relevant methods, such as statistics. Some rules of thumb and “folklore” of the GNSS industry are presented, especially when the practicalities of receiver design is at odds with the theoretical considerations.

This chapter is divided into four subsections, which concentrate on the following topics:

- *Signal processing issues that need to be taken into account in difficult environments:* To get more sensitivity, longer integration times are needed in acquisition and tracking. Because the receiver will encounter both strong and very weak signals, it needs to cope with possible cross-correlations between signals from different satellites, as well as interference from other sources. In difficult locations such as urban canyons, the signals are disturbed by multipath effects, and it is desirable to employ mitigation methods to reduce their influence.
- *Aiding and assistance possibilities and systems that can be used to support GNSS:* Assistance data can be provided to the receiver to narrow down the

search window in acquisition and thus to help weaker signals to be acquired in a reasonable time. Providing initial position and the navigation message content as assistance data eliminates the need for data decoding in the receiver and allows navigation with much weaker signals. There are also several possible supporting systems for GNSS, which provide the receiver with some additional measurements that can be used to increase the positioning availability.

- *Special computational aspects and filtering techniques for weak and noisy signals:* Since in most applications the receiver movements are quite limited, the navigation algorithms can utilize information on user behavior. The reduction of the degrees of freedom in the navigation solution (by assuming the receiver is on or very near the Earth, for example) can allow algorithms to cope with fewer measurements, thereby increasing the positioning availability. Weighting schemes can be used to improve the receiver's performance in varying signal environments. In the case of road applications, the positioning can be constrained by map matching. Since algorithms that are optimized for very poor signal environments are not optimal in good signal conditions, the receiver will benefit from adaptiveness in the algorithms.
- *Fault detection and exclusion methods for indoor and weak environments:* In such environments where the measurements are noisy and disturbed by multipath and interference, it is especially important to be careful with the measurement quality control. Traditional measurement quality control methods are concerned mainly on rare failures of the GPS satellites, mainly because these methods were developed originally in the context of aviation (as described in Chapters 4 and 10). In noisy environments, the algorithms need to pay more attention to the possibility of frequent and multiple failures in the measurements. Navigation quality control is applied to estimate the reliability of the navigation output, and to reject the result in case the quality is too low compared to a predefined criterion.

12.2 Signal Processing Considerations Related to Weak Signals

This section presents some signal processing issues related to weak signals. Our focus is mainly on consumer-grade devices, which normally use only the GPS C/A code. References to GPS in this section thus apply mainly to the C/A code and the L1 frequency. Nominally, according to the GPS ICD specification [3] the GPS C/A code signal at L1 is designed to arrive on the ground in good LOS conditions at a power level of approximately -158.5 dBW. However, obstructions often attenuate the GPS signal, as the signal reflects from buildings and propagates through them. Indoors, building materials such as glass, concrete, steel, insulation, drywall, wood, and other materials further attenuate the signal. Some models suggest that attenuation inside buildings can be in the order of 2.9 dB per meter of structure [4]. However, the obstructing material plays a key role in the amount of attenuation: metals are heavily obstructing whereas nonconducting materials typically attenuate less. Outdoors, tree foliage also attenuates the signal, making dense forests some-

what challenging environments. Additionally, the power level may decrease due to fading caused by interference from reflected signals.

The tracking of GPS signals requires that the receiver generate a PRN replica for the satellite to be tracked and correlate it with the incoming signal. The correlation is often done with an early-minus late DLL, and the carrier phase or frequency is tracked with a Costas or FLL. These are described in detail in Chapter 2 and in [5, 6] and demonstrated in the software receiver included in Chapter 5. For weak signals and signals that are corrupted by effects such as multipath, the simplest signal processing methods are not sufficient. Here are some of the most important things that must be taken into account in a weak signal environment:

- *Longer integration:* In the case of weak signals, the integration times in acquisition and tracking need to be longer. This leads to longer acquisition times, emphasizing the need to reduce the search windows by using assistance information about the carrier frequency and code phase. In tracking, the longer integration times also lead to reduced dynamics in the loop. To get the most sensitivity, it would be preferable to increase the coherent integration times, but this is limited by the data bit length, unless the data content is available as assistance and can be removed from the signal.
- *Cross-correlation:* Due to the rather small cross-correlation threshold of the GPS L1 signal (21 dB), any strong GPS signals easily cross-correlate with other PRNs. This can be seen as perturbations in the weaker signal or even as situations where the receiver tracks a completely nonexistent signal.
- *Multipath:* Especially in urban situations, reflections of signals are a major source of errors in the GPS measurements.
- *Hardware issues:* When designing hardware for receivers that are intended for weak signal conditions, it is important to minimize sensitivity losses in the hardware. Thus, the receiver design should make sure that, for example, components are properly shielded to avoid interference and leakage and that the oscillators and other analogue components are of good enough quality. The details of the GNSS chipset used should also be analyzed carefully to assess the suitability of the hardware for high sensitivity applications.

Table 12.1 describes signal levels typical for different environments. The SNRs given are only rough estimates, and in a particular receiver these levels can differ by several decibels.

The aspects of GNSS signal processing that require extra care when weak signals are considered are acquisition, fine acquisition, bit synchronization and data detection, code and carrier tracking, and navigation message decoding. Special algorithms need to be considered when in addition to the weak signals there are also strong interfering signals present. Usually, one attempts to acquire the strong signals first, where they may be removed from the processing to acquire robustly the remaining weak signals [4]. Comprehensive discussions on GNSS receiver algorithms for weak signals can be found in [4, 7, 8]. In addition, [9] discusses achieving high detection sensitivity utilizing extremely long coherent integrations.

Table 12.1 Effect of Signal Level on GPS Receiver

<i>Approximate SNR</i>	<i>Description of Effects in GPS at This Signal Level</i>
−130 dBm	Nominal signal level; acquisition can be made without integrating over multiple code epochs.
−140 dBm	A Costas loop has difficulties following the signal; an FLL should be used instead of a Costas loop below this level.
−142 dBm	There are so many bit errors that it becomes difficult to extract navigation data.
−145 dBm	The integration times needed for acquisition are so long that it is not practical to do an unassisted acquisition.
−150 dBm	A signal at nominal level creates cross-correlations at about −150 dBm. If a receiver tries to handle signals below this at all, it will need cross-correlation mitigation methods.
−155 dBm	The DLL and FLL have difficulties; below this, advanced tracking algorithms are needed.
−165 dBm	Weakest signals trackable with currently available stand-alone commercial receivers.
−170 dBm	To track signals at this level and below, the integration times are so long that high-precision clock oscillators need to be used in the receiver.

12.2.1 Acquisition of Weak Signals

In order for a GNSS satellite signal to be acquired, the receiver must find the correct frequency and code phase for the signal. There are many variations on the details of the acquisition, but typically the acquisition proceeds by first defining a coherent and noncoherent integration time to be used, as well as a set of carrier frequency bins. The signal is multiplied by the carrier replica and correlated with differently phased code replicas. Typically all code phases need to be considered, so that to obtain a half-chip resolution, a receiver would need to correlate 2×1023 differently phased code replicas, in the case of the GPS C/A code. Many GPS receivers have a dedicated search engine to generate these code replicas. Alternatively, receivers can speed up the correlation somewhat by using FFT techniques, which have proven to be especially useful in software-based receivers. After correlation, the signals can be accumulated according to the coherent integration time, after which the power for each code phase is calculated. The signals can then also be accumulated according to the noncoherent integration time. The whole process is performed for each frequency bin. Finally, the acquisition decision is made by comparing the strongest peaks in the accumulated results against a set of acquisition thresholds. The acquisition process produces an approximate Doppler shift and an approximate code delay of a satellite signal. The precision of these values depends on the number of frequency bins and the number of code phases used in the acquisition. Usually, the accuracy of the estimation is not suitable to initialize the signal tracking, requiring a fine acquisition step before tracking can be initiated (see Chapter 5 or [7]).

The acquisition of very weak signals requires long coherent and incoherent integrations. The effect of the coherent and noncoherent integration times on the power is discussed in Example 1 in Section 12.6.1, which illustrates this effect using simulated GPS data. Coherent integration decreases the noise bandwidth and improves sensitivity faster than noncoherent integration, but increasing the coherent integration time also reduces the effective width of the frequency search bin,

requiring more frequencies to be searched. This may make the search too time-consuming unless frequency assistance is available. In addition, the coherent integration is sensitive to the signal phase, and therefore it is limited by the data bits and bit edges, which are unknown at the acquisition time. Thus, a single coherent integration of a GPS signal can last only 20 ms. This duration can be increased by using knowledge of the individual data bits. By inverting the data when the data bit sign changes, the effect of the navigation data can be removed and the coherent integration time can be made longer. This process is usually called data wiping or bit wiping. The data bit information can be provided to the receiver as assistance in applications where a suitable server is available. Also, some bits of the navigation message remain constant or predictable and can also be used for data wiping. The TLM and HOW words of each subframe contain up to 60 bits that are constant or can be predicted from the time of week, and can therefore be used to obtain quite long coherent integrations.

Theoretically, by extending the integration times, it would be possible to acquire arbitrarily weak signals, assuming that the receiver remains stationary. However, in reality the quality of the receiver's own clock oscillator places a limit on the integration times, as discussed in Section 12.2.2.

In weak signal environments, the results of the acquisition are very noisy and easily affected by interference from other signals, as demonstrated by Example 1 in Section 12.6.1. When the integration times become longer and the found signals become weaker, it becomes increasingly difficult to make correct acquisition decisions and to distinguish real signals from false acquisitions caused by noise and interfering signals. When acquiring weak signals, there will always be a number of false acquisitions that will end up in the tracker, and they will typically be dropped after some time. Depending on the details of the tracking and acquisition architecture, the integration times and acquisition thresholds need to be adjusted to obtain the desired sensitivity with an acceptable amount of false acquisitions. This will be mainly an experimental task, since a reliable theoretical analysis would need to take into account the details of the acquisition and tracking process of the receiver's software architecture, which would be a very complex task.

12.2.2 Clock Stability and Integration Times

As the signals get weaker, longer integration times are needed in both acquisition and tracking. Assuming that the data bits are removed from the signal, the coherent integration time can be increased beyond 20 ms. Beyond 20 ms, the integration times will be limited mainly by the receiver clock oscillator stability. The short- and long-term stability of clock oscillators is a rather complicated topic, and the behavior of oscillators depends on many effects. An introduction can be found in [10]. Below we will discuss only a very rough estimate of how the clock stability limits the integration times.

Generally, for a coherent integration to be useful, the carrier of the incoming GNSS signal and the replica signal generated from the receiver oscillator should stay in the same phase relative to each other during the whole integration period. If the signals drift more than a small fraction of the carrier period, the integration result will be partially cancelled out. Since the carrier frequency is about 1.5 GHz,

the period is about $1/1.5 \text{ GHz} = 0.6 \text{ ns}$. Thus, during a coherent integration the clock should stay within 0.1 ns or so. The oscillators typically used in commercial receivers are temperature-controlled crystal oscillators (TCXOs) with short-term stabilities from 10^{-8} to 10^{-9} [10]. From this we could estimate that with a TCXO, one should not expect to use coherent integration times longer than roughly $0.1 \text{ ns}/10^{-9} = 0.1 \text{ s}$.

If longer coherent integration times are needed, more stable clocks such as oven-controlled crystal oscillators (OCXOs) can be used. They can have short-term stabilities of up to 10^{-12} [10], which by the same reasoning would allow integration times of several tens of seconds. If very long integration times are used, it will be necessary to take into account the changing frequency of the incoming GNSS signal due to satellite accelerations and other effects. A discussion of very long coherent integrations can be found in [9].

The maximum noncoherent integration time can be estimated in a similar way. During a noncoherent integration the change in carrier phase can be ignored, but the PRN codes in the incoming signal and the replica need to stay within a fraction of a chip from each other. In the case of GPS signals, the clock should not drift much more than about $0.1 \mu\text{s}$. From this, we could estimate that noncoherent integrations of a few tens of seconds should be possible with a TCXO. With an OCXO, very long noncoherent integrations indeed should be possible, but with such long times, our simple estimate based on the short-term stability figure of 10^{-12} is likely not to be sufficient.

12.2.3 Tracking of Weak Signals

When the signals get weaker, the noise variance of the tracking loops needs to be reduced by using longer integration times. This increases the mean time to lose lock. To be able to benefit as much as possible from coherent integrations, it is necessary to synchronize the integration to the bit boundaries, which enables the use of coherent integrations up to one data bit long (20 ms in GPS C/A code). Data wiping can also be used in tracking to increase the coherent integration times, up to lengths that are feasible with the TCXO used in the receiver. Integration times longer than this need to be done noncoherently.

A receiver's software will typically contain separate tracking loops for the code and carrier. When the signals get weaker, the carrier PLL, or Costas loop, is in practice the first one to drop. FLLs can handle weaker signals, as can the code DLL. Because PLLs provide much more reliable Doppler data than FLLs, it is desirable to have some loop switching, where the carrier tracking is done with an FLL below a certain signal level.

The tracking loops also need to be adjusted to make their bandwidth narrower. The narrower tracking loops make the receiver less sensitive to the dynamics in the data, meaning that the receiver will drop signals more easily when the receiver is accelerating. This may not be practical from the application point of view. Therefore, it is almost essential to have at least some adaptiveness in the loops: For strong signals, wider loop bandwidths must be used.

Data decoding from weak signals is very difficult since there are significantly more bit errors when the signals are weak. The parity algorithm used in the GPS

signal is not very efficient, and though it can theoretically be used to correct some bit errors, this is not very useful in practice. Typically, the bit error rate grows very rapidly once the signal level goes below a certain SNR (see Table 12.1 for the approximate level), which makes it very difficult to improve the data decoding sensitivity by such means. In practical circumstances the SNR is not constant but fluctuates according to the receiver movements. In such cases it may be useful to decode individual words from the GPS data and to extract useful data from snippets shorter than 6s [11]. In addition, using long-term or extended ephemerides as described in Section 12.3.1 is very useful.

12.2.4 Cross-Correlation and Interfering Signals

A receiver that is capable of handling weak signals will necessarily be more sensitive also to interference caused by other GNSS signals or by signals from other sources than GNSS. If such signals are wrongly interpreted by the receiver as GNSS signals, then they will obviously affect the positioning solution considerably. Interference can come from many sources, including electronic and industrial equipment. This section concentrates on interference from cross-correlating strong GNSS signals. In addition, the effects of cross-correlations are illustrated by Example 1 in Section 12.6.1.

The cross-correlation threshold of the Gold codes used in the GPS signal is 21 dB. From the acquisition and tracking point of view this means that a signal whose SNR is X dBHz, can be seen as a “ghost” for another PRN, with an SNR of $X-21$ dBHz. The cross-correlated signal does not differ particularly from a real signal, so it is easy to mistakenly use it in navigation, and since the measurements correspond to a wrong PRN, this may lead to very large navigation errors. In addition, since the GPS signal does not include the PRN number as part of the navigation message, it is possible for the receiver to start decoding data from a cross-correlated signal. If a complete ephemeris is decoded from such a signal and used in navigation, it will lead to completely arbitrary navigation results.

To cope with cross-correlated signals, it is necessary to use the quality and integrity monitoring methods described in Section 12.5. In addition, mitigation methods specific to cross-correlation can be used. A simple mitigation method to cope with cross-correlation relies on disallowing signals that are weaker than the cross-correlation threshold. In theory, a threshold of “strongest signal—21 dB” would be sufficient. In practice, a higher threshold should be used because several strong PRNs can contribute to a cross-correlation at the same frequency. To allow this kind of PRN selection, the strong signals should be acquired first, then the weaker ones. This can be done by performing two acquisitions: one with short integration times to find all the strong signals, then another round with long integration times for those PRNs that were not found in the first round.

A more complicated mitigation technique that is also applicable to non-GNSS signals is interference cancellation, in which the interfering signal is identified and subtracted from the incoming signal. After this is done, the acquisition process and tracker can then operate on cleaner data. In the case of cross-correlating GNSS signals, the strong signal can be acquired first, and then subtracted from the signal before acquiring and tracking the weaker signals [12, 13].

12.2.5 Multipath Mitigation

One of the major problems in urban areas is that the signals are reflected from buildings, with these reflections causing distortions in the received signal. Outdoors, the multipath phenomenon most often encountered is that both the direct signal and its reflection are received, with the reflected signal usually much weaker in power. Sometimes, it can also happen that the reflection exceeds the direct signal in power or that the direct signal disappears altogether (echo-only signal). Indoors, the situation is usually much more complex than outdoors since the direct signal is very weak and there are typically many reflecting signals. Some outdoor environments are also extremely challenging, such as streets between very tall buildings, which sometimes provide good reflection surfaces. With only multipath signals present in poor-signal environments, acquisition and tracking schemes face difficult obstacles and some error is bound to remain in the resulting measurements.

Depending on the carrier phases of the reflected and LOS signals, they can interfere constructively or nonconstructively, which leads to a change in the signal level and to a deformation in the form of the autocorrelation peak, as illustrated in Figure 12.1. The deformed autocorrelation peak will lead to an offset in the code phase measurements. It is also possible that the LOS signal is obstructed completely and only the reflected signals are present. These multipath effects on the code phase measurements are most prominent for reflections whose delay is less than one chip, meaning that the effects of it on the code measurements are typically up to a few tens of meters, a couple of hundred at most.

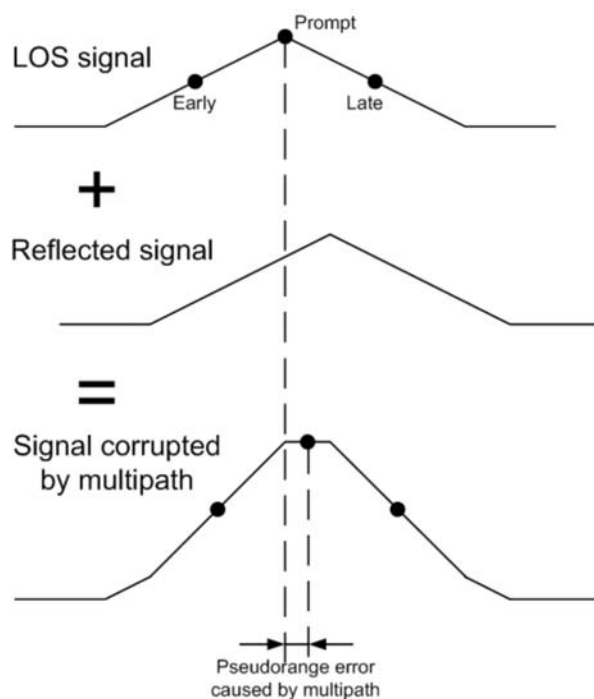


Figure 12.1 Illustration of how the correlation peak is affected by a multipath signal and how that affects the code phase measurement from the tracking loop.

Because of the short carrier wavelength of the signal (about 20 cm), the interference can change from constructive to nonconstructive when the receiver is moving, or when the geometry changes slowly due to the satellite movements. Often seen symptoms of this in a multipath environment are rapid changes in signal level for a moving receiver and “oscillations” in the signal level for a stationary receiver. As an example, Figure 12.2 shows the SNRs for three satellites from a stationary receiver. Two of the three satellites are clearly affected by multipath.

See, for example, [14] for an analysis of measurements to characterize the typical behavior of multipath signals in an indoor environment. Additionally, descriptions of a number of multipath mitigation methods can be found in [15]. Multipath mitigation is an active research topic, and many different mitigation methods have been proposed. The following is a list of some methods in this domain.

- Tracking methods that are less sensitive to typical multipath errors, such as the narrow correlator, which was one of the first multipath mitigation methods proposed [16], and the strobe correlator.
- Methods that try to explicitly estimate the different multipath components, and then use those to estimate the phase of the probable LOS signal (e.g., the multipath-estimating DLL) [15].
- Using antennas that are less sensitive to multipath. Several types of antennas have characteristics that reduce multipath disturbances. Choke-ring antennas are sensitive mainly to signals coming from above, and therefore are less sensitive to multipath disturbances coming from the ground. These antennas

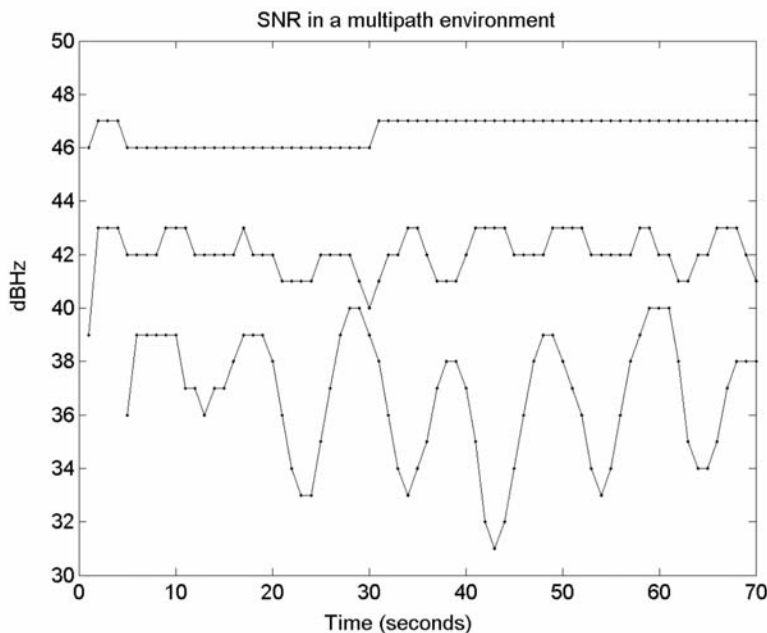


Figure 12.2 Oscillations in the SNR from a stationary receiver in a multipath environment.

are less attractive for a pedestrian receiver, because of size and price. Circular polarized antennas also reduce multipath, because the GPS signal is circularly polarized, and its polarity inverts when it is reflected from a surface. Therefore, a circularly polarized antenna diminishes multipath signals that are reflected only once. In high-end systems, it is also possible to use adaptive arrays of antennas, which can be made sensitive only to signals coming from the directions of the satellites. But their high cost prevents their usage in low-cost devices.

12.2.6 Benefits of Future GNSS

The future GNSS currently under development will provide many benefits for navigation with weak signals. Here, we consider some points relevant to the Galileo system, but similar considerations apply to other systems as well (e.g., modernized GPS signals):

- The Galileo PRN codes are longer and have better cross-correlation properties than GPS C/A code signals.
- The Galileo signals will use MBOC modulation, which provides possibilities for better multipath mitigation. However, the MBOC modulation scheme will require more complicated tracking methods to prevent the receiver from tracking a wrong correlation peak.
- For a receiver that can handle Galileo and GPS signals, there will be approximately twice as many satellites visible at one time, which will increase positioning availability.
- The better error correcting codes in the Galileo signal will allow data bits to be decoded at lower signal levels.
- A dataless pilot signal will be included, allowing long coherent integration times to be used even without the navigation data. This will allow acquisition of weaker signals.
- The signals will be transmitted at a slightly higher power than GPS C/A code signals.
- The signals' data will include an identifier of the transmitting satellite, which will better enable the detection of cross-correlating signals.

12.3 Aiding Possibilities and Supportive Systems

Often stand-alone GNSS are not sufficient enough in severely obstructed signal conditions for providing means to obtain a location. Assistance is required in order for a GNSS receiver to be able to acquire and track the satellite signals. In addition, supportive systems for backing up GNSS or as additional support and verification are essential, especially in indoor conditions.

12.3.1 Assistance

Because of the long integration times needed for acquisition and the difficulties in decoding the navigation message, successful weak signal positioning is very

dependent on outside assistance information. There are three main assistance ingredients to acquiring the signals and computing the positioning result: the satellite orbits, time (which is needed to calculate the satellite position from the orbit info), and receiver location (which is needed to predict the satellite Dopplers in acquisition).

A-GNSS can be employed within a wireless network. The network can send the GPS receiver, for example, a prediction of the received signal including Doppler shift, code delay, initial position estimate, and the navigation message [17, 18]. This enables the use of a long coherent integration time enhancing sensitivity and reducing the search space of the acquisition process.

Normal ephemerides are valid up to 4 or 8 hours, and their precision degrades quite fast after that, though they can still be used for navigation in some circumstances. For an analysis of the degradation of ephemeris and almanac precision, see [19]. It is also possible to use an orbit propagator to predict the satellite positions quite reliably beyond the 4- or 8-hour ephemeris fit interval. Predictions can remain usable for several days and can reduce the need to decode navigation data in the receiver. Several variations of such extended ephemeris solutions have been proposed in articles and in commercial products. More information on long-term ephemeris can be found in [20].

The actual information fields of A-GNSS are included in several standards today, including those of 3GPP and OMA (Open Mobile Alliance). Notably, minimum operational performance values are defined for both mobile station-based A-GPS and mobile station-assisted A-GPS. A comprehensive summary of A-GNSS and its standardization can be found in [21].

12.3.2 Supportive Systems for GNSS

Many methods have been proposed to support GNSS in weak signal conditions with additional sources of measurements that can be used for positioning. As also discussed in Chapter 9, these include:

- Pseudolite infrastructures in which a set of “pseudo-satellites” or pseudolites send a GNSS signal that is tracked by the receiver similarly to a normal GNSS satellite [22, 23]. The pseudolites need to be installed specifically for GNSS support, which can be quite costly, but on the other hand the signals can be tracked and used by ordinary GNSS receiver hardware. The receiver software needs to be adapted, because the navigation data at least in the GPS system does not allow for a stationary transmitter. The main technical difficulties in pseudolite systems arise from the fact that the receiver is close to the transmitter, and therefore when the receiver moves, the pseudolite signal can become strong enough to mask the real GNSS signals. Because of this “near/far” problem, the pseudolite signals are only usable within a relatively narrow zone around the transmitter antenna, unless mitigation methods such as a suitable pulsing of the signal are applied. Pseudolites are often used in airports, where they can be used to provide integrity information in addition to additional positioning signals. The Galileo system includes

plans for specifying a local component, which will specify the use of pseudolites and other local augmentation systems.

- MEMS sensors for dead-reckoning assistance [24, 25]. The use of sensor data allows the receiver to determine its own movement regardless of the GNSS signals. In research literature, the integration of sensor data to GPS navigation has concentrated mainly on high-end systems, but since sensor prices have decreased and the need for consumer grade devices has increased, there has recently been a growing interest towards very low-cost MEMS sensors for GNSS support. A thorough discussion of the computations for a full inertial navigation system (INS) can be found in [26], which is targeted mainly for high-accuracy inertial navigation but is also applicable to MEMS INS. Also, see Chapters 6 and 7 for more information on GPS/INS. There have also been useful methods published where only a small number of sensors are used instead of a complete MEMS inertial navigation solution. Examples include step-counting measurements for pedestrian applications, odometer measurements for vehicle applications, air pressure sensors for altitude measurements, and magnetic sensors as low cost substitutes for gyros.
- WLAN systems [28–30] that are today occupying more and more public areas. In WLAN systems, which are basically cellular networks consisting of several cells each served by a base station [29], the measurements most often are based on the signal levels of the WLAN signals. A map of the signal levels in the area is built beforehand, and this map is used to match the receiver's position. This prominent indoor positioning methodology is called WLAN fingerprinting. Accuracies in the range of a few meters can be achieved with WLAN fingerprinting. However, WLAN fingerprinting requires an on-line “training” period for making the necessary measurements to create the radio map. The map may become obsolete if there are changes in the WLAN infrastructure or the environment. In addition, the WLAN operational range is fairly limited.
- Cellular positioning [17, 28, 29, 31] is a significant means of providing GNSS support, either in GNSS position initialization, hybrid positioning for different measurement combinations, or as a backup system. Methods such as cell-ID, E-OTD, and TDOA are well-known [29].
- Digital TV broadcasts are based on accurately time-stamped signals that can be used to get pseudorange measurements to the broadcasting antennas. Details can be found in [28, 32].
- RFID technology is based on radio signals that are exchanged between an RFID reader and an RFID tag. The range of the signals is quite short, and therefore RFID tags and readers can be used to implement proximity sensors [29]. Bluetooth signals can be used similarly.

Figure 12.3, adapted from [17], illustrates how different positioning technologies differ in their accuracy and their availability in different environments. Figure 12.3 also illustrates the amount of specific infrastructure that would be needed to obtain a good coverage with that technology. No single technology exists that would provide accurate positioning in all environments. Therefore, it is clear that

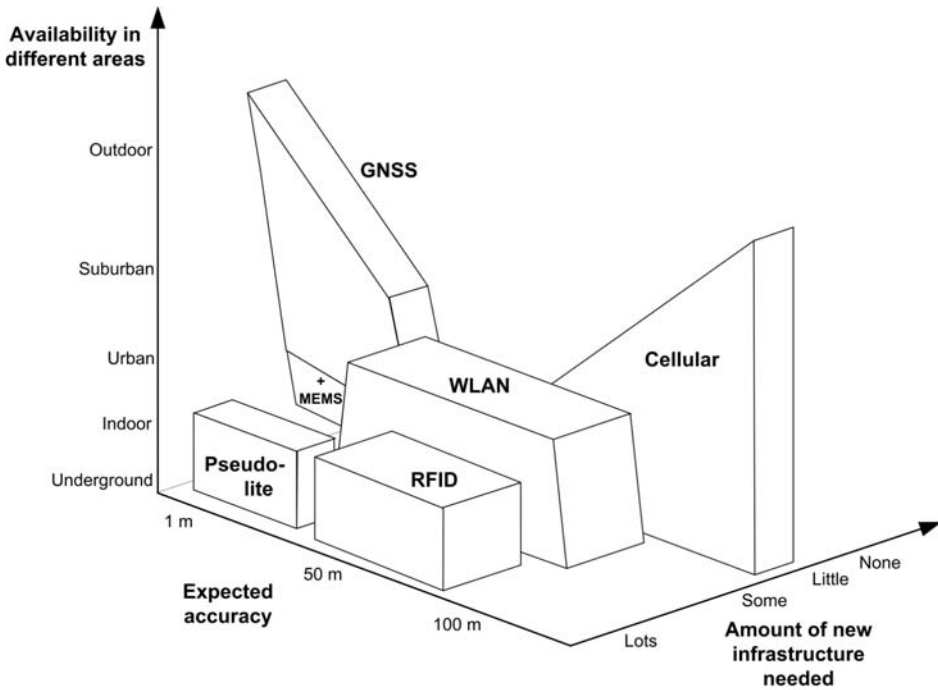


Figure 12.3 Comparison of positioning technologies in terms of accuracy, availability, and infrastructure needed.

to provide full positioning service coverage accurately in all environments, several technologies should be combined.

12.4 Navigation Algorithms for Difficult Signal Conditions

The previous sections of this chapter have concentrated on signal processing algorithms specific to difficult conditions and assistance methods that may be used to overcome some of the problems in these conditions. This section outlines some methods that can be applied when the navigation solution is computed, such as using a priori knowledge of the user application, map-matching, and adapting to the changing environment.

Difficult signal conditions in urban areas and inside buildings suffer from poor GNSS signal quality and lack of signals. GNSS navigation accuracy depends both on the measurement errors and the signal reception geometry. Poor satellite geometry amplifies the effect of the measurement errors in the navigation result. Another difficulty in these environments is that the signal quality can change very quickly when the receiver is moved. Therefore, signals often get so weak that they are dropped from tracking and need to be reacquired. Reacquisition of a weak signal is much more difficult than keeping track of a signal that gradually deteriorates.

Because of the poor and changing signal availability, quality, and geometry, the navigation continuity also suffers. Thus, the application often cannot provide continuous navigation accuracy and integrity throughout the operation. This

navigation noncontinuity can be a problem for some applications, such as mapping and routing algorithms.

As an example of how the navigation quality is affected indoors, Figure 12.4 presents a position scatter from an indoor environment, compared to the same results from an open-sky environment. The plots show the deviation in the east and north directions from the true position. The results were gathered with a commercial high-sensitivity GPS receiver with state-of-the art positioning performance. The indoor data set included four GPS signals per epoch on average, while the outdoor data set had 10 measurements per epoch on average. The position solution availability was 92% in the indoor case (close to a window inside an office building) and 100% in the outdoor case. The indoor result illustrates typical behavior caused by scattered signals, poor geometry, and attenuation.

12.4.1 Constraints on User Motion

In most applications, the real degrees of freedom in the receiver movements are quite limited. For example, in a typical consumer (e.g., automotive) application the movement occurs mainly horizontally, along roads, and the accelerations are rather small. Taking these restrictions properly into account will improve the navigation results especially in weak signal conditions since it allows positioning with fewer measurements, and provides more redundancy for quality control.

Moreover, often additional information on the user is available (e.g., the user’s altitude or velocity). Usage of this additional information will result in an increased measurement redundancy for the solution computation as well as the quality check-

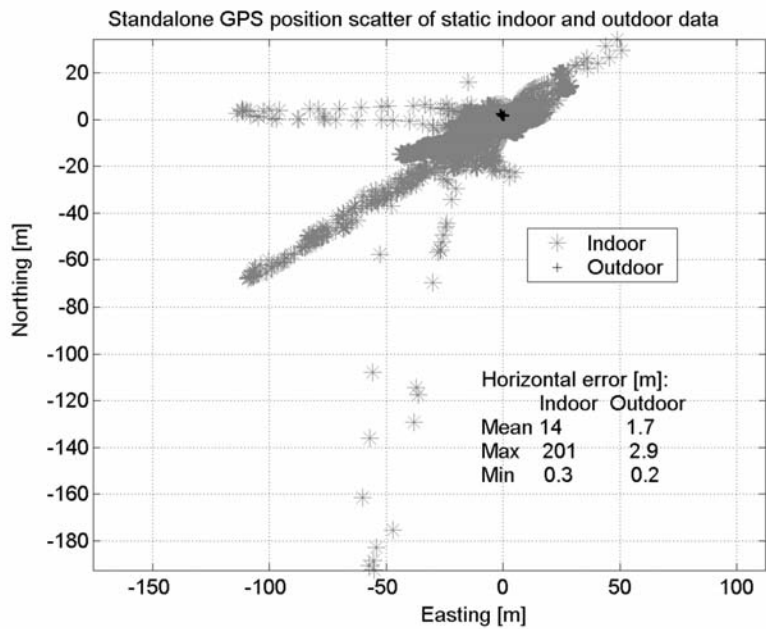


Figure 12.4 Scatter results of indoor and outdoor GPS data with typical parameter estimation techniques.

ing. This information can be incorporated strictly as a condition in the quality checking; however, to account for the uncertainty of this additional information it is often better to treat it as an additional measurement [33]. If the variance of this additional observation is low with respect to that of the pseudorange or pseudorange rate measurements, the observation acts like a constraint.

It is also possible to take the constraint into account when computing the navigation solution. For example, if it is known that the user moves mainly in a horizontal direction, then it is possible to compute the navigation solution in a two-dimensional horizontal plane. Then one measurement less is needed and navigation availability can be improved. Or, it may be known that the user moves predominantly in relatively straight curves (for example, a car, or as a more extreme example, a train). Such a constraint could be incorporated into the navigation Kalman filter, with the drawback of making the filter more complex. Similar to the quality checking, the constraints can be introduced either in the receiver movement model (i.e., in the Kalman filter's state update matrix), or it can be provided as an additional aiding measurement. In the case of mainly horizontal movement, the constant altitude can be provided with the aid of a pseudorange measurement from a fictive satellite placed at the center of the Earth, as discussed in [34].

12.4.2 Map Matching

Map matching is especially useful in urban areas and applications such as automotive navigation, where the user moves only on roads. Roads and streets are provided to the map-matching software as a database, and the map-matching algorithm finds a likely position on the streets that corresponds to the position calculated by the navigation algorithm. In urban areas, the GPS errors resulting from multipath and cross-correlation are often so large that a naive matching method where the position is matched to the nearest road is not sufficient. Instead, a sophisticated matching algorithm must take into account also the recent history of receiver movements. In such algorithms, care must also be taken to recalculate the matching when it is found that the matched position is in error. Many algorithms have been proposed for map matching, from rather simple heuristic models to more sophisticated methods such as Kalman and particle filtering. For a review of the current state of research, see [35]. For improved matched position output, it would be more effective to couple the map matching directly in the navigation filter, since in effect navigation on a street is a one-dimensional problem and a position could be calculated from only two pseudorange measurements. However, since traditionally the GPS navigation software and the map-matching software have been built by different companies, such coupling would be difficult to implement. Therefore, a typical automobile navigation device calculates its map-matching results by using the position and velocity from the GPS navigation solution.

12.4.3 Adaptive Algorithms

In good visibility conditions, a good navigation solution can be obtained by simple parameter estimation techniques such as least-squares methods. In good conditions it is not absolutely necessary for example to adapt the measurement weighting to

estimates of the signal quality. However, in weak signal positioning, where the poor quality signals have large errors, it is necessary to implement proper measurement variance modeling and related weighting. The measurement noise estimate should be adapted to the variable-power observations. Most of the time, good results can be obtained by describing the measurement uncertainty with a simple signal-to-noise-based model [36].

In weak-signal conditions, where the operating environment is often quickly varying, it is important that the algorithms for computing the navigation solution can adapt to the changing situations. It may also be desirable to make the constraining algorithms adaptive, so that when the signals are weak and there are few measurements, the movement is constrained more, and in good signal conditions, more freedom is allowed. The thresholds in different parts of the positioning method and quality assurance approaches need to be adaptable depending on the severity of the environment degradation. In addition, techniques such as multiple-model Kalman filters [37] can be used to calculate the positioning output with several differently configured filters and to choose the one that most closely matches the measurements.

12.5 Quality and Integrity Monitoring

This section discusses the importance and methodology of navigation result quality and integrity monitoring in weak-signal environments for accuracy and reliability assurance. Most of the traditional integrity-monitoring literature is based on the assumption that at most one measurement is incorrect. This assumption is too limiting in difficult signal environments, so we also discuss ways to treat multiple errors, in addition to giving an outline of traditional integrity monitoring techniques.

12.5.1 Introduction to Integrity Monitoring

Integrity is often defined as the ability of the navigation system to provide timely warnings to the user when it is inadvisable to use the system for navigation. The GPS system currently provides some basic integrity information to users via the navigation message, but is not timely enough for safety-critical applications. Integrity monitoring in the receiver is essential in safety-critical applications in order to ensure a certain predetermined degree of integrity for the navigation function during some mode of operation. A typical example of an application where integrity monitoring is essential is aviation. However, integrity monitoring techniques are also important in less critical applications to ensure a user position solution with an acceptable quality even with weak signals. RAIM [38–43] is a means of providing integrity with the capability of detecting when a satellite failure or a measurement error has occurred. RAIM is based on checking the solution consistency and requires redundant measurements: In three-dimensional positioning where at least four measurements are needed for computing the position and time, fault detection requires at least five measurements and fault exclusion at least six. In the case of a constrained navigation solution, for example with constant altitude, then in

principle four and five measurements are enough for detection and exclusion, respectively.

Traditionally, RAIM and its performance have mostly been associated with the integrity-monitoring tasks in aviation and other safety-critical applications, where relatively good LOS signal reception conditions prevail, and there are often strict rules of well-defined integrity modes. The goal in these traditional RAIM operating environments has been to eliminate large measurement errors due to, for example, satellite clock or ephemeris failures, which can be caused by failures in the satellite or in the control segment. These errors are very rare, with a typical rate of one error per 18 to 24 months in the GPS system. Due to the rarity of these errors, it is usually assumed in traditional RAIM literature that only one erroneous measurement—also called a blunder or an outlier—needs to be detected in one measurement epoch. Integrity monitoring is discussed in more detail in Chapter 4 (DGNSS) and specifically for aviation in Chapter 10.

The GNSS measurements can also be erroneous for reasons other than control or space segment problems. In indoor and other weak-signal environments, the most significant errors are caused by multipath propagation and attenuation. In such poor signal environments, also the measurements with a high SNR are often erroneous because of multipath effects. Signals with a lower SNR can be of good quality, only attenuated by obstacles, but they can also be completely incorrect, for example when the receiver has locked into a completely false signal (cross-correlation). Because of such effects, weak signal conditions result in a large number of erroneous measurements, and RAIM is an important function for assuring good accuracy in weak signal conditions in any application.

12.5.2 Reliability Testing

Positioning with GNSS is based on computing the user position from measurements of the satellite-to-user ranges. Because of the difference in the receiver and satellite clocks, the measurements are inevitably biased, and thus the measurements are called pseudoranges, denoted by ρ . GPS pseudorange positioning is discussed in detail in [5, 6, 8, 44–46], as well as in Chapter 3. For simplicity in the presentation of this chapter on reliability testing, we present a snapshot approach where we calculate a weighted least-squares solution on one measurement epoch, as opposed to a solution where previous epochs are also taken into account. In addition, we concentrate only on the computation of position from the pseudoranges. Computing the velocity from Doppler measurements proceeds in an almost identical manner, as shown in Chapter 3, and the development of the RAIM/reliability testing techniques for velocity is identical.

For the basic analysis described below, it is assumed that the errors in the measurements are normally distributed and that they are uncorrelated. Additionally, it is assumed that there will be only one erroneous measurement, an outlier, in one epoch as this significantly simplifies the presentation. The implications of these assumptions will be discussed after presenting the simpler case. An outlier is often defined as an observation that appears to be inconsistent with the remainder of the data and the assumptions made, and thus deviates significantly from the other

observations, as discussed in [47]. Figure 12.5 shows outliers appearing among GPS range measurements.

12.5.3 Weighted Least-Squares Notation

The vector of the reduced pseudorange measurements, $\Delta\boldsymbol{\rho}$, is obtained as

$$\Delta\boldsymbol{\rho} = \begin{bmatrix} \delta\rho^1 \\ \vdots \\ \delta\rho^k \end{bmatrix} = \boldsymbol{\rho}_c - \boldsymbol{\rho}_0 \quad (12.1)$$

where

$$\boldsymbol{\rho}_c = \begin{bmatrix} \rho_c^1 \\ \vdots \\ \rho_c^k \end{bmatrix}$$

are the measured pseudoranges that have been corrected by the atmospheric corrections, and

$$\boldsymbol{\rho}_0 = \begin{bmatrix} \rho_0^1 \\ \vdots \\ \rho_0^k \end{bmatrix}$$

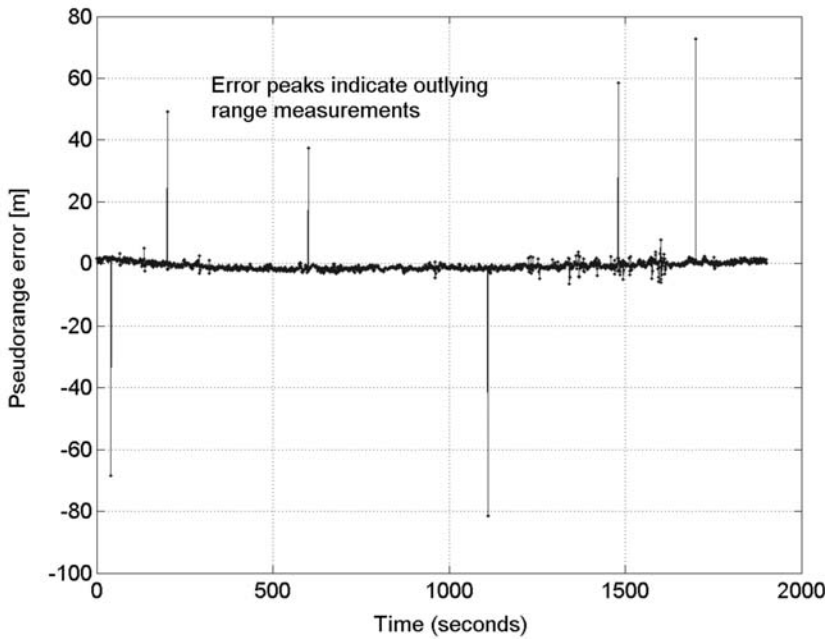


Figure 12.5 Outliers in GPS range measurements.

are the predicted pseudoranges. Equation (12.1) follows the general method of the derivation starting at (3.6a), which expresses the reduced pseudorange measurement (i.e., the measured minus predicted pseudorange, for the k th satellite). The predicted pseudoranges ρ_0 are computed with the aid of the linearization point, \mathbf{x}_0 . If the linearization point is sufficiently close to the true user position, the estimation problem can be represented as a linear equation

$$\Delta \rho = \mathbf{G} \Delta \mathbf{x} + \boldsymbol{\epsilon} \quad (12.2)$$

where \mathbf{G} is the *design* or *geometry matrix* calculated using the direction vectors to the satellites, and $\boldsymbol{\epsilon}$ is the vector containing the pseudorange measurement errors. For details, see (3.6a)–(3.8b) and [5, 6, 44]. Equation (12.2) is similar to (3.8a). The vector $\Delta \mathbf{x} = [\delta \mathbf{x} \ \delta b]$ is the vector offset of the user's true position and time from the linearization point. Since $\Delta \rho$ is contaminated by the unknown random errors in $\boldsymbol{\epsilon}$, (12.2) should be treated as a stochastic equation and $\Delta \mathbf{x}$ should be determined using parameter estimation techniques. A weighted least-squares (WLS) estimate is a best linear unbiased estimate, yielding the lowest estimation error among all linear estimators [48], when the weight matrix \mathbf{W} is obtained from the variance covariance matrix $\boldsymbol{\Sigma}$ [see (12.4)]. With the assumption that $\mathbf{G}^T \boldsymbol{\Sigma}^{-1} \mathbf{G}$ is nonsingular and that there are at least as many independent observations as unknowns, the WLS estimate of $\Delta \mathbf{x}$ is

$$\Delta \hat{\mathbf{x}} = (\mathbf{G}^T \mathbf{W} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{W} \Delta \rho = (\mathbf{G}^T \boldsymbol{\Sigma}^{-1} \mathbf{G})^{-1} \mathbf{G}^T \boldsymbol{\Sigma}^{-1} \Delta \rho \quad (12.3)$$

The WLS estimate of (12.3) is comparable to the least squares equation of (3.10). The only difference is the weighting of the pseudorange measurements based on their relative noise levels and relative importance to each estimated quantity [5]. The $\mathbf{G}^T \boldsymbol{\Sigma}^{-1} \mathbf{G}$ quantity is nonsingular when $\boldsymbol{\Sigma}$ is nonsingular, which in turn is true when the direction vectors to the satellites are linearly independent. The weight matrix \mathbf{W} can be expressed as

$$\mathbf{W} = \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_k^2 \end{bmatrix}^{-1} \quad (12.4)$$

where σ_k^2 is the variance of the k th pseudorange measurement.

As discussed in Section 12.4.3, the measurement uncertainties vary in the inconsistent environments of weak signal conditions. The variance estimates for each observation of the measurement noise levels can be obtained, for example, from an approximation of the pseudorange error variance versus satellite elevation angle [5], or versus the SNR [5, 33]. A simple C/N_0 -based model (carrier-to-noise ratio), originally discussed in [49], for the pseudorange error variance can be of the form

$$\sigma_k^2 = a + b * 10^{-(C/N_0)/10} \quad (12.5)$$

where the constants a and b should be chosen according to the environment and equipment and the C/N_0 is given in units of decibel-Hertz (dBHz). In [36], values of $a = 10 \text{ m}^2$ and $b = 150^2 \text{ m}^2\text{Hz}$ are suggested.

The estimate of the unknown user coordinates is obtained by correcting the linearization point by the estimate of the error $\Delta\hat{\mathbf{x}}$, as also presented in (3.11a),

$$\hat{\mathbf{x}} = \mathbf{x}_0 + \Delta\hat{\mathbf{x}} \quad (12.6)$$

To cover the case where the linearization point is further away and to still make sure that the linearization error in (12.2) is negligible, the same estimation process is usually repeated by taking the corrected output as a new linearization point and repeating until the sequence of estimation points converges.

12.5.4 Residuals and Redundancy

In case the estimation problem is overdetermined (i.e., when there are more independent measurements than unknowns), the measurements are said to be redundant. Since the measurements are affected by noise and other errors, the measurement vector will not fit exactly the WLS solution, and there will be some pseudorange residuals left. These residuals are defined as the difference between the estimated values of the observations and their corresponding measured values [50]. The residuals are calculated as,

$$\hat{\mathbf{v}} = \mathbf{G}\Delta\hat{\mathbf{x}} - \Delta\boldsymbol{\rho} = -\mathbf{R}\Delta\boldsymbol{\rho} \quad (12.7)$$

where \mathbf{R} is a projector from the reduced observations to the weighted least-squares residuals. If the measurement errors are zero-mean normally distributed with a variance $\boldsymbol{\Sigma}$, i.e. $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, then it can be shown that the residuals are distributed as $\hat{\mathbf{v}} \sim N(\mathbf{0}, \mathbf{R}\boldsymbol{\Sigma}\mathbf{R}^T)$. These residuals can be used to test the internal consistency of the measurements [51], as will be described in Sections 12.5.5 and 12.5.6. Assuming that the measurement model is correct, the observational residuals indicate the extent to which the measurements agree with each other.

From (12.7) and (12.3), one can derive the equation

$$\mathbf{R} = \mathbf{C}_{\hat{\mathbf{v}}}\boldsymbol{\Sigma}^{-1} \quad (12.8)$$

where the matrix $\mathbf{C}_{\hat{\mathbf{v}}}$ is the variance covariance matrix of the residuals, defined in (12.10). The trace of the matrix \mathbf{R} is the overall redundancy i.e. the degree of freedom [33, 52, 53] and therefore \mathbf{R} is referred to as the *redundancy matrix*. With uncorrelated observations, this matrix plays a key role in quality control because it allows us to determine if there is enough redundancy for reliability testing. To detect and identify m outliers, the overall redundancy has to be at least $m + 1$.

The i th diagonal element of matrix \mathbf{R} , denoted by r_i , corresponds to the contribution of the i th observation to the overall redundancy, but it is also the scale factor by which a bias in the i th observation will affect the residual. In some measurement geometries, it can happen that r_i is close to zero, and then the i th

observation contributes very little to the redundancy. This is not desirable because then errors in that measurement cannot be detected. A zero redundancy number implies a completely uncontrolled observation.

The residual vector can also be used to check for modeling errors, as indicated in [50]. However, the process is more complicated, and usually it is assumed that the model is correct.

The process of outlier detection starts by first performing a *global test* to detect the presence of outliers, and then a *local test* to identify the largest outlier which is then rejected. The process may be iterated for several outliers. Sections 12.5.5 and 12.5.6 detail these steps. The process includes assuming zero-mean measurement errors and uncorrelated observations. The implications of correlated measurements are briefly discussed later in Section 12.5.12.

12.5.5 Global Test

The global test for assessing whether the set of measurements includes errors or not is performed by calculating first a test statistic $T = \hat{\mathbf{v}}^T \boldsymbol{\Sigma}^{-1} \hat{\mathbf{v}}$. It can be shown [54] that when the measurement errors are zero-mean normally distributed [i.e., $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma})$] the test statistic $\hat{\mathbf{v}}^T \boldsymbol{\Sigma}^{-1} \hat{\mathbf{v}}$ follows a central chi-square distribution. In the case where there are bias errors (i.e., when the mean value of the measurement errors is not zero) the test statistic follows a noncentral chi-square distribution whose noncentrality parameter (λ) depends on the magnitude of the error [36, 51, 53]. See Figure 12.6 for an illustration of these distributions.

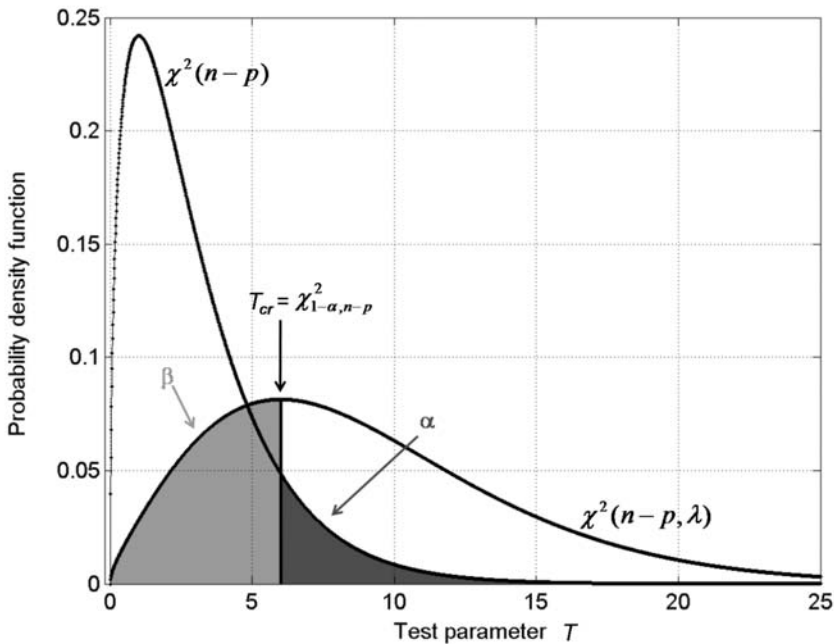


Figure 12.6 PDFs of the chi-square distributions related to the global test ($n-p = 3$, $\alpha = 0.11$, $\beta = 0.2$, $\lambda = 6.03$).

To execute the test, we need the critical value T_{cr} of the test. This is the threshold against which the test parameter $T = \hat{\mathbf{v}}^T \mathbf{\Sigma}^{-1} \hat{\mathbf{v}}$ is compared. If $T > T_{cr}$, an inconsistency is assumed and the local test is performed to identify which measurement is the cause of the inconsistency.

The critical value is chosen so that the confidence level is $1 - \alpha$. The parameter α is a significance level that determines the “tightness” or “looseness” of the global test (i.e., the probability of false alarm). The false alarm probability is the probability at which the test statistic T still exceeds the critical value, even when it is centrally chi-square-distributed. In case of errors in the measurements, the test statistic will follow a noncentral chi-square distribution and the test statistic T will exceed the critical value by a larger probability. Given the false alarm probability and the degree of freedom of the chi-square distribution, the critical value can be obtained from the cumulative distribution function of the chi-square distribution, denoted by $T_{cr} = \chi_{1-\alpha, n-p}^2$. The value $n-p$ is the degree of freedom for the distribution, where n is the number of measurements available, and p is the number of parameters to be estimated. In the case of 3-D navigation, $p = 4$, and in the case of 2-D navigation with constant altitude, $p = 3$.

A second parameter, a probability β of accepting erroneous measurements is also involved in the global test as well as a noncentrality parameter λ . The central and noncentral chi-square probability distribution functions and probabilities related to the global test are illustrated in Figure 12.6. The criterion for selecting the critical value for the global test is that the cumulative distribution function of the central chi-square distribution has a value $1 - \alpha$ while the cumulative distribution function of the noncentral chi-square distribution has a value β , i.e., $\chi_{1-\alpha, n-p}^2 = \chi_{\beta, n-p, \lambda}^2$. Section 12.5.8 gives more details and explanations on setting the detection threshold (i.e., the critical value T_{cr}) and on the derivation of the noncentrality parameter.

12.5.6 Local Test

When the global test fails, we assume for now that the reason is a single outlying observation. We postpone the discussion of multiple outliers until Section 12.5.9. With only one outlier, testing is easy, and outlier identification can be attempted if the redundancy is at least two.

The pseudorange residuals, the vector $\hat{\mathbf{v}}$, can be standardized by their own variance, resulting in the variable [55]

$$w_i = \frac{\hat{v}_i}{\sqrt{(\mathbf{C}_{\hat{\mathbf{v}}})_{ii}}}, \quad i = 1, \dots, n \quad (12.9)$$

which is assumed to be normally distributed with a zero mean and a variance of one [i.e., $w_i \sim N(0, 1)$]. In (12.9), n denotes the number of observations and the matrix $\mathbf{C}_{\hat{\mathbf{v}}}$ is the covariance matrix of the residuals mentioned above,

$$\mathbf{C}_{\hat{\mathbf{v}}} = \mathbf{\Sigma} - \mathbf{G}(\mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G})^{-1} \mathbf{G}^T \quad (12.10)$$

The standardized residuals w_i can be easily used for outlier detection since they are normally distributed, and if the reality does not meet this assumption an error is detected.

In the local test, the i th measurement is assumed to be correct when $-T_{cr} \leq w_i \leq T_{cr}$, where T_{cr} is the critical value of the local test. The local testing is thus performed by taking the absolute value of the standardized residual as the test statistic, $T = |w_i|$. The critical value is calculated from the false alarm probability α_0 . Because $w_i \sim N(0, 1)$, T_{cr} can be obtained from the inverse of the cumulative distribution function of the normal distribution with probability $1 - \alpha_0/2$, denoted as $n_{1-(\alpha_0/2)}$, shortened hereafter as $T_{cr} = n_{1-(\alpha_0/2)}$.

In summary, the local testing is thus based on the comparison

$$|w_i| \leq n_{1-(\alpha_0/2)} \quad (12.11)$$

We use $\alpha_0/2$ here because the false alarm can occur at both the positive and negative values of w_i . For the local test, the false alarm rate α_0 is predetermined and it is illustrated in Figure 12.7 along with the zero-mean and biased PDFs of the normal distribution related to the local test.

The assumption was that the measurement set contains at most one erroneous measurement. Therefore the local test is done only when the global test fails (i.e., has indicated the presence of an outlier), and only the observation with the largest value of $|w_i|$ is tested. If the local test detects an erroneous measurement, then this measurement is rejected and the remaining measurements are used to calculate the

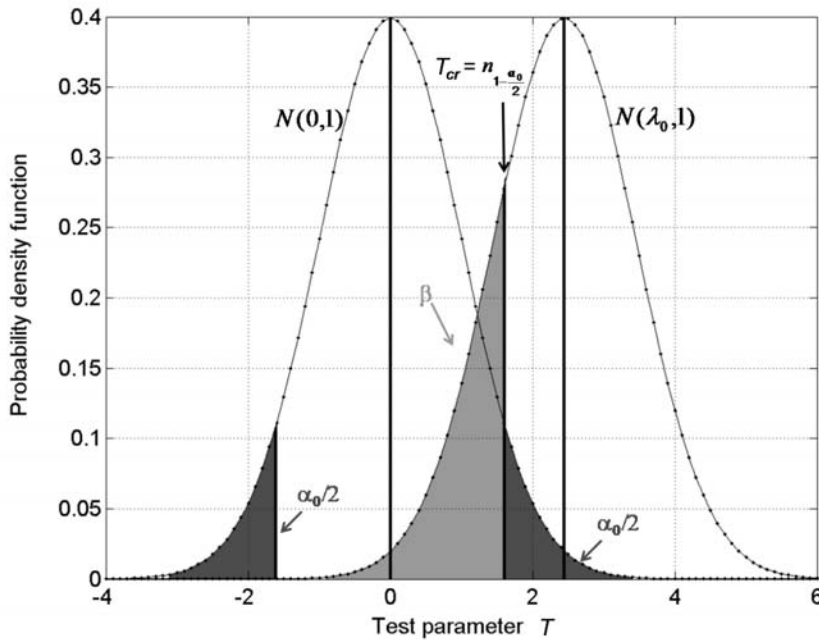


Figure 12.7 PDFs of the normal distributions related to the local test ($\alpha_0 = 0.107$, $\beta = 0.2$, $\lambda_0 = 2.45$).

solution. If the global test fails, but the local test does not indicate an erroneous measurement in the observation set, the solution is concluded to be unreliable.

The reliability testing or outlier detection procedure used in the local test was introduced originally by Baarda in 1968 [55] and is also known as data snooping.

12.5.7 Null Hypothesis and Alternative Hypothesis

It is customary in statistics to formulate tests using the *null* and *alternative* hypotheses. This formulation is useful in isolating the conditions that will result in incorrect decisions by the quality-control procedure. In the case of the global test described above, the null hypothesis H_0 states that the model is correct and the measurements agree with the probability distribution that was assumed [i.e., that the measurement errors are distributed as $\epsilon \sim N(0, \Sigma)$ and there are no bias errors in the set of measurements]. The alternative hypothesis H_a states that the model is wrong, or that the assumption is not correct, which means that there are at least some bias errors in the set of measurements and the errors are not zero-mean but rather $\epsilon \sim N(\mu, \Sigma)$. More details on these techniques can be found in [55, 56]. In the case of the local test, the null hypothesis of no error, $w_i \sim N(0, 1)$, is denoted by $H_{0,k}$ and the alternative hypothesis of a biased measurement, $w_i \sim N(\lambda_0, 1)$, by $H_{a,k}$. When wrong decisions are made, these are often classified as a type I or type II failure. This hypothesis testing formulation is summarized in Table 12.2 together with the associated probabilities.

12.5.8 Parameters for Fault Detection and Exclusion

The fault detection and exclusion equations can be summarized as follows.

Global test for fault detection:

$$H_0 \text{ (no failure): } \hat{\mathbf{v}}^T \Sigma^{-1} \hat{\mathbf{v}} \leq \chi_{1-\alpha, n-p}^2 \quad (12.12)$$

$$H_a \text{ (integrity failure, i.e., something is wrong):} \quad (12.13)$$

$$\hat{\mathbf{v}}^T \Sigma^{-1} \hat{\mathbf{v}} > \chi_{1-\alpha, n-p}^2$$

Table 12.2 Hypothesis Testing Table

	<i>Decision</i>	
	<i>H₀ is accepted</i> (all measurements are used)	<i>H₀ is rejected</i> (some measurements are rejected)
<i>Situation</i>		
<i>H₀ is true</i> (all measurements are good)	Right decision Probability $1 - \alpha =$ confidence level	Type I failure (rejecting good measurements) Probability $\alpha =$ significance level
<i>H₀ is false</i> (some measurements are erroneous)	Type II failure (accepting erroneous measurements) Probability β	Right decision Probability $1 - \beta =$ power of the test

Local test for fault exclusion:

$$H_{0,k} \text{ (measurement } k \text{ is not erroneous): } |w_k| \leq n_{1-(\alpha_0/2)} \quad (12.14)$$

$$H_{a,k} \text{ (measurement } k \text{ is erroneous):} \quad (12.15)$$

$$|w_k| \geq |w_i| \forall i \wedge |w_k| > n_{1-(\alpha_0/2)}$$

For a consistent set of decisions, the false alarm rate of the global test, α , must relate to that of the local test, α_0 , together with the probability β of missed detection, which is the same for both the global and the local tests [55, 57]. An erroneous measurement that causes a global test failure should be indicated by the local test with the same probability. Thus, only two of the values α , α_0 , and β can be chosen arbitrarily. It is customary to first select the false alarm probability α_0 , then β , and then solve for the global test false alarm probability α . The values of these parameters are linked by the following equations:

$$\lambda = \lambda_0^2 = (n_{1-(\alpha_0/2)} + n_{1-\beta})^2 \quad (12.16)$$

$$\chi^2_{(\beta, n-p, \lambda)} = \chi^2_{(1-\alpha, n-p)} \quad (12.17)$$

where λ is the noncentrality parameter of the noncentral chi-square distribution related to the global hypothesis testing and λ_0 is the expected value of the normal distribution (the bias in case of an error) related to the local test [33]. These parameters and the probability functions are illustrated in Figures 12.6 and 12.7.

12.5.9 Multiple Outliers

In reality, the assumption of a single outlier is a severe restriction. It has been found that data snooping can also cope with multiple blunders if it is performed iteratively [36, 47, 58]. In this case, the observation with the largest absolute value of w_i is excluded and the whole process of parameter estimation, statistical tests in Equations (12.12) and (12.14), and possibly the rejection of an observation are repeated, until no more outliers can be identified.

If m outliers are suspected, a redundancy of at least $m + 1$ is needed in order to identify them. However, due to the mutual influence of observations (i.e., an error of one observation is absorbed by the residuals of all observations) erroneous rejection of a good observation is possible, especially with large or multiple biases [59]. In degraded signal environments, the redundancy is generally poor and thus it is desired to keep as many observations as possible for obtaining an efficient estimate [33, 36, 60]. Therefore, if more than one observation is being excluded, the iterated reliability checking should include a reconsideration of observations rejected earlier [61]. Figure 12.8 presents the simplest form of an iterative fault detection and exclusion procedure, but it can be modified in several ways according to the application at hand.

The iterative method described works well in practice and does not consume significant computing resources. Also other methods for multiple measurement

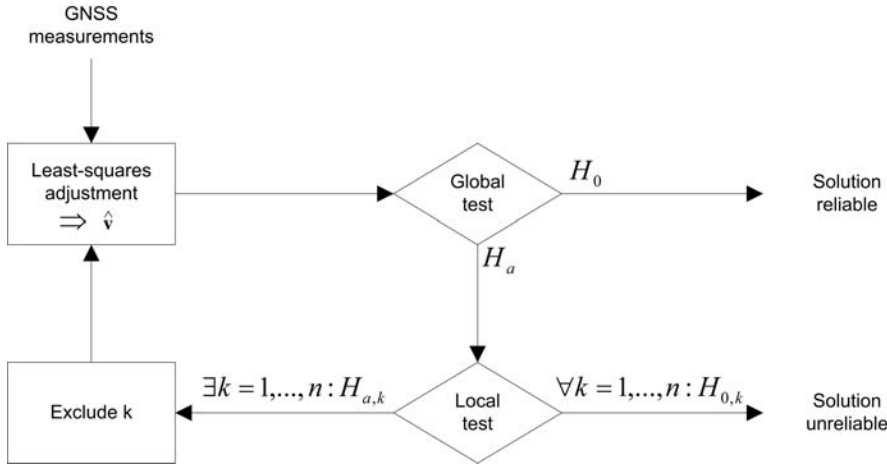


Figure 12.8 Simple iterative fault detection and exclusion procedure.

rejection have been proposed, such as “subset testing” discussed in [36], which considers all possible subsets of measurements and can thereby detect and more reliably identify multiple errors. Development of RAIM with protection for multiple biases is discussed comprehensively in [62] in a context of aviation protection level service coverage. The drawback of most proposed methods for multiple failure protection is that they are computationally much more intensive than simple data snooping and yet do not provide dramatic performance improvements. Multiple measurement rejection therefore remains an active research topic in GNSS RAIM literature, especially associated to applications where certain protection levels need to be achieved.

12.5.10 Fault Detection and Exclusion in Kalman Filtering

For simplicity, we have presented the RAIM techniques in the context of snapshot WLS navigation solutions, where the history of the measurements is not handled. For better navigation output, most receivers use a Kalman filter–based solution, for example as described in Chapter 6 and [63]. The same strategies of reliability testing presented here can be applied to Kalman filtering [37] as well. In this case, the global and local tests need to be based on the innovations as discussed in [64].

In Kalman filtering the possibility for a modeling error is typically much more pronounced than in single-point least-squares estimation. This is because the Kalman filter equations must include a model of the receiver dynamics. Therefore, it is important that the quality-control process takes into account the possibility of an incorrect model or divergence of the Kalman filter. These topics are outside the scope of this discussion, but are well presented in Kalman filtering textbooks, for example [37, 63].

12.5.11 Quality Control

In the overall quality control of a navigation solution, the geometrical considerations should be taken into account as well. The quality of the user position estimate

depends not only on the quality of the range measurements but also on the user/satellite observation geometry. The DOP, which is a simple quality measure of the user/satellite geometry as described in Chapter 3 and [5, 6, 44] should not surpass a threshold even after rejection of outliers. A poor geometry may amplify random errors and biases and therefore contribute to large position errors, as illustrated in Figure 12.9.

An overall quality control procedure in weak signal environments should include assessment of the user redundancy, the solution consistency, and the user-satellite geometry. A navigation solution should be considered reliable only if there is enough redundancy to perform reliability testing, all observations are controllable (no zero redundancy numbers), consistency checks are successful, and solution geometry is acceptable. Figure 12.10 presents a simple scheme for total solution quality control.

12.5.12 The Practical Side of Quality Control

The preceding discussion on quality monitoring is based on the assumption that the measurements are uncorrelated and that the measurement errors are normally distributed. In practice, especially in weak signal environments, this may not be true. Many factors will apply that are difficult to analyze theoretically, and therefore a receiver will need to contain a significant amount of heuristic decision logic to cope with invalid measurements. For example, the errors in measurement data that are caused by multipath effects are often correlated in time and not normally distributed. A typical multipath effect can grow slowly and increase up to several tens of meters. In addition, multipath errors can be correlated between different satellites. In practice, it is not rare to see a situation in which most of the channels are affected by slowly varying errors. Because of such effects, the statistical tests described in this section may not perform well. It is often useful to devise heuristic

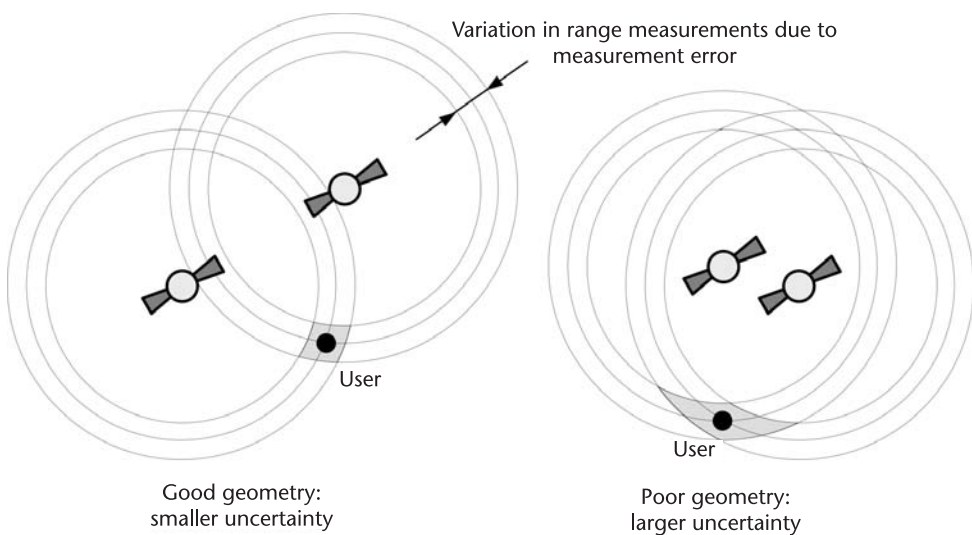


Figure 12.9 Poor satellite geometry strengthens the effect of the measurement errors.

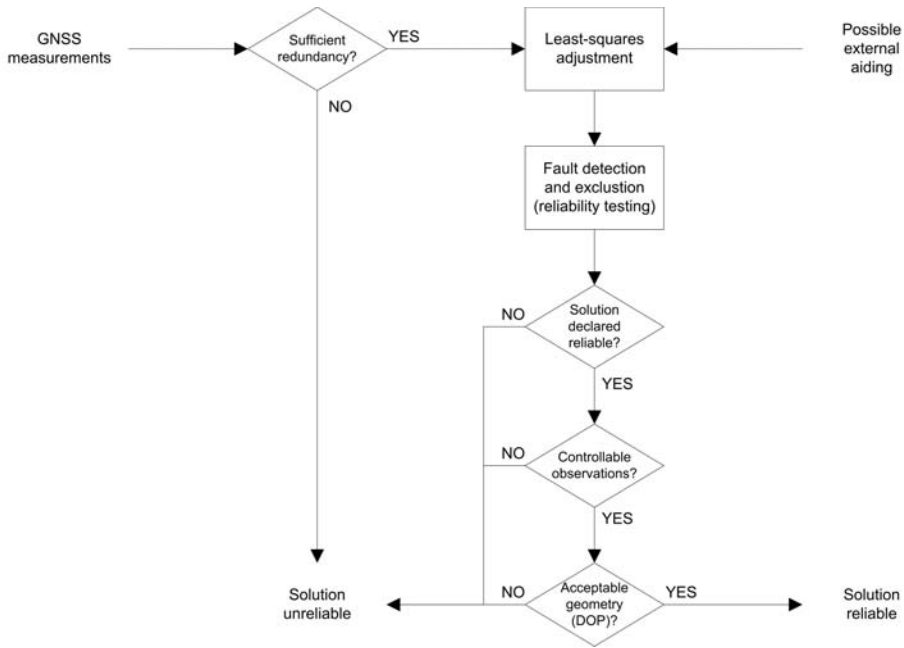


Figure 12.10 Navigation solution quality control.

tests in addition to the statistical tests. For example, after analyzing the behavior of the data of the particular receiver, it is possible to devise additional tests based on criteria such as the SNR, number of measurements, and magnitudes of the residuals.

Should the observations be correlated and the resulting covariance matrix is a full matrix, then (12.10) is still the covariance matrix of the residuals but the parameter w_i in (12.9) is no longer a suitable indicator for outliers [33, 56, 57].

Errors such as divergence of a Kalman filter can also be detected with the statistical tests outlined above. However, in such cases most of the residuals are affected by an erroneous linearization point, and the statistical tests often cannot correctly identify the measurements that are erroneous. It is therefore useful to include some heuristic tests to detect divergence and other problems in the model. If an erroneous model is detected, it is often best to disable the navigation outputs until the navigation calculations can be reinitialized with better data.

It should be noted that the statistical tests will deem a solution unreliable if it is not redundant. In practice this would mean that solutions with only four satellites (three in two-dimensional positioning) would be rejected. In many applications it is preferable to accept such solutions, because otherwise availability may be too low. For example, it may be desirable to assume that measurements with high SNR are reliable, and to bypass the statistical tests for nonredundant solutions that have been computed from strong signals. When intensive algorithms exceed the available CPU budget, it may be more efficient to detect faulty measurements with simple heuristic rules instead, such as rejection based on SNR or the magnitude of the innovations in a Kalman-filter based navigation solution.

12.6 Examples Included on the DVD

This section contains examples on MATLAB/Octave scripts, which illustrate some of the issues relevant to weak signal processing. The first example in Section 12.6.1 demonstrates the difficulties of acquiring weak signals, and the second example in Section 12.6.2 presents the outlier detection algorithms discussed in Section 12.5.

12.6.1 Example 1: Acquisition of Weak Signals

This example is intended to illustrate some effects that pose difficulties when acquiring weak signals, namely the noisiness of the integrated data and cross-correlation effects. To run the example, run the script `example1_main.m` in MATLAB or Octave. The execution of the script may take some time. The script will produce plots similar to the ones shown below and a few additional ones, but since the samples are generated randomly, the detailed contents of the plots will change between each run. To get a feel for the typical behavior, you may wish to execute several times. In addition, it is a simple matter to change items in the scripts; please refer to the scripts and the comments there.

The script mentioned above uses a couple of auxiliary scripts:

- The `generate_iq.m` script is used to generate a sequence of simulated samples. The signal is simulated according to a very simple model, with no data bits, constant frequencies, and some other simplifying assumptions, which are described in the comments of the script. Although the signal simulation here is very simple and ignores many receiver-dependent details, the results shown in the plots below are also seen in real receivers.
- The `search.m` script implements the basics of a simple search algorithm. It generates a replica signal and correlates that with the signal. The correlation results are output for plotting.
- The `ca_code.m` script generates the PRN sequences.

Table 12.3 lists PRNs based on simulated data that is 100-ms-long, with a sampling rate of about 4 MHz. The SNRs are given as decibels below the noise level. The strongest signal, PRN 1, corresponds approximately to a GPS signal at the nominal level, as described in Section 8.4 of [44].

Figures 12.11 and 12.12 show the result of searching for PRN 1 by correlating only 1 ms of data. Figure 12.11 shows the maximum value of each frequency bin and Figure 12.12 shows the correlation result for each code phase in the frequency bin that has the largest maximum. In Figure 12.12, there is a clear peak at the correct code phase of 100 chips, but it is not very large compared to the noise.

Table 12.3 PRNs Included in the Simulated Data

<i>PRN</i>	<i>SNR</i>	<i>Frequency</i>	<i>Code Phase</i>
1	−18 dB	0 Hz	100 chips
2	−28 dB	700 Hz	200 chips
3	−38 dB	70 Hz	300 chips

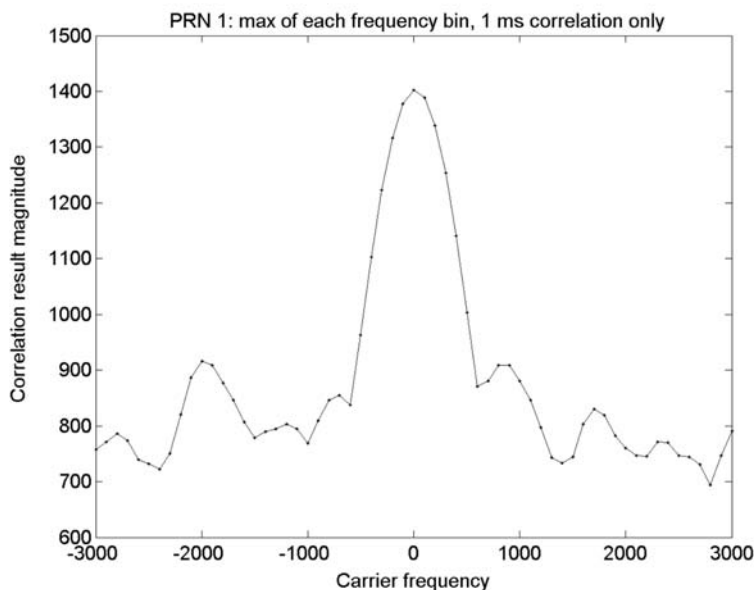


Figure 12.11 Maximum of each frequency bin for PRN 1.

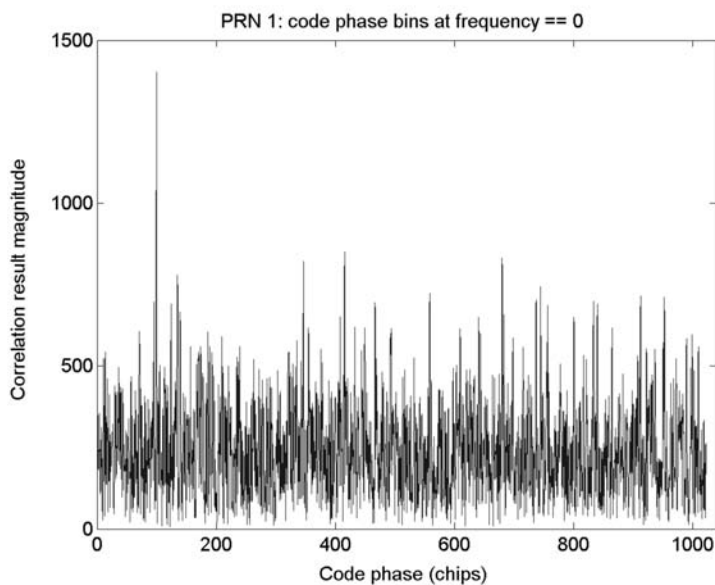


Figure 12.12 Correlation result for each code phase in the frequency bin with the largest maximum.

Indeed, when repeating the script it does happen that the noise peaks are larger than the correct peak, which would lead to an incorrect acquisition.

When searching for the weakest signal, which is 20 dB lower than the strongest one, we would expect that the cross-correlation from the strong signal would be noticeable. To find the weakest signal, we need to integrate longer. In this example, we used a coherent integration time of 20 ms, and five noncoherent integrations.

The results are shown in Figure 12.13. There is a clear maximum at the correct 70 Hz frequency, but also a clear peak at 0 Hz, which is caused by the cross-correlation from PRN 1. In some runs of the script, the cross-correlation peak is even larger than the correct one. Figures 12.14 and 12.15 show the code phase results at frequencies 70 Hz and 0 Hz, respectively. The correct peak shows up

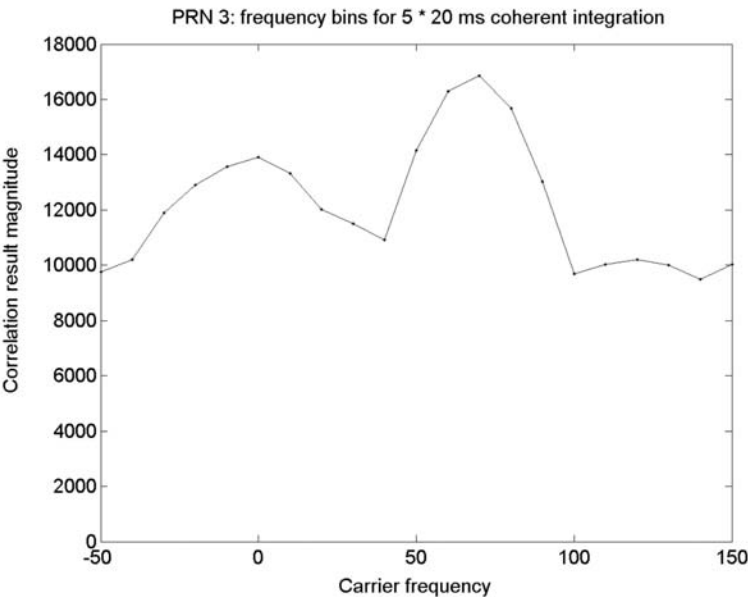


Figure 12.13 Frequency bin plot showing PRN 3 and cross-correlation from PRN 1.

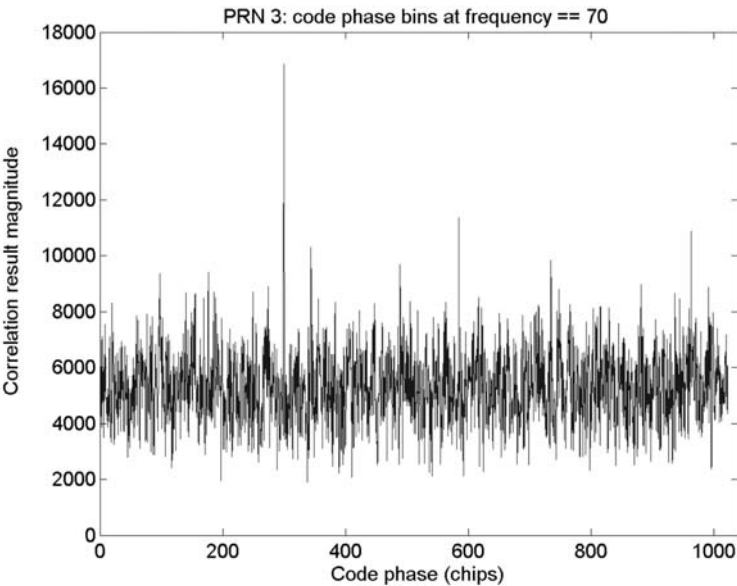


Figure 12.14 Code phase plot at 70 Hz, showing the correct code phase.

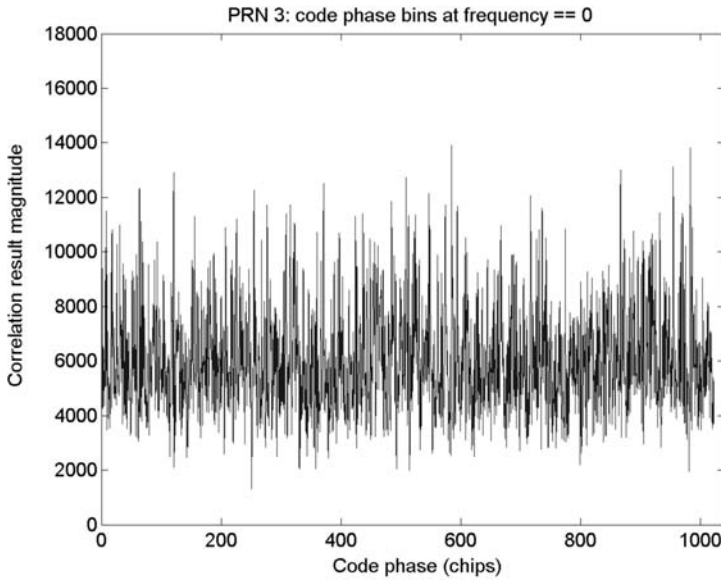


Figure 12.15 Code phase plot at 0 Hz, showing many cross-correlation peaks.

clearly in the 70-Hz frequency bin, whereas in the 0-Hz frequency bin we see a large number of peaks that are larger than the background noise, and that are almost as strong as the correct peak at 70 Hz. If the weak signal would be at the same frequency as the strong one, the cross-correlation peaks could easily cause an incorrect acquisition result. Another problem is that in reality, the frequencies change during tracking. If the receiver is tracking a weak signal and the cross-correlation peaks of a strong signal move so that they coincide with the correct peak, the tracking loops may be tricked into tracking the wrong signal unless cross-correlation mitigation methods are used.

12.6.2 Example 2: Fault Detection and Exclusion

This example illustrates how measurement errors can be detected and excluded in order to improve the position solution quality. A file `roof_data.mat` is provided for loading to the Matlab or Octave workspace. It includes a structure with GPS satellite pseudorange measurements and corresponding satellite coordinates from a real-life static scenario with induced pseudorange errors (similar to the illustration in Figure 12.5). The main example script, `example2_main.m`, produces a WLS position solution from these measurements and includes applying iteratively the global test for fault detection and the local test for fault exclusion until the WLS solution is found consistent. Finally, the example script produces figures illustrating the position results.

The main script uses a couple of auxiliary scripts:

- The script `positioncalc.m` produces the actual WLS result.

- The script `raim.m` performs the global and local tests depending on the inputs given. The probabilities are set using the method described in Section 12.5.8 ($\alpha_0 = 0.107$, $\beta = 0.2$, and $\alpha = 0.11$).
- Scripts `wgsxyz2lla.m`, `wgsxyz2enu.m`, and `wgslla2xyz.m` are related to coordinate transformations necessary for position solution illustration purposes.
- The script `calcchisquare.m` provides the detection threshold for the global test.

Figure 12.16 shows the WLS position result for the given measurement data in a local coordinate frame (northing-easting) when all the available measurements are used, and highlights the positions that have been detected to be erroneous by the global test.

The script outputs the measurements detected as outliers by indicating the faulty PRN. Figure 12.17 presents the positioning result when these faulty observations have been excluded from the WLS solution and illustrates the significant improvement in positioning accuracy.

12.7 Summary

Successful weak-signal GNSS positioning requires different kinds of techniques for aiding and monitoring the navigation computation. Compared to traditional GNSS techniques typical for good signal conditions, additional processing power and memory are unavoidably consumed in order to enhance the navigation reliability in obstructed environments. Methods both in the core signal processing (acquisition,

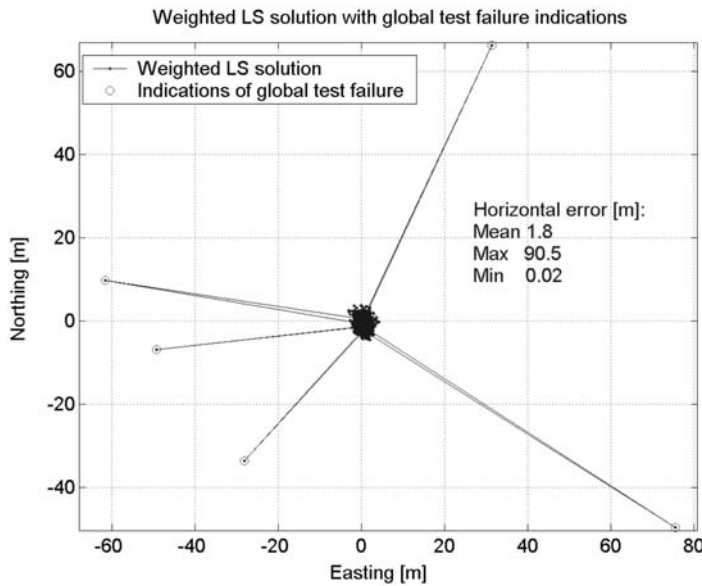


Figure 12.16 The WLS positioning result with all available measurements.

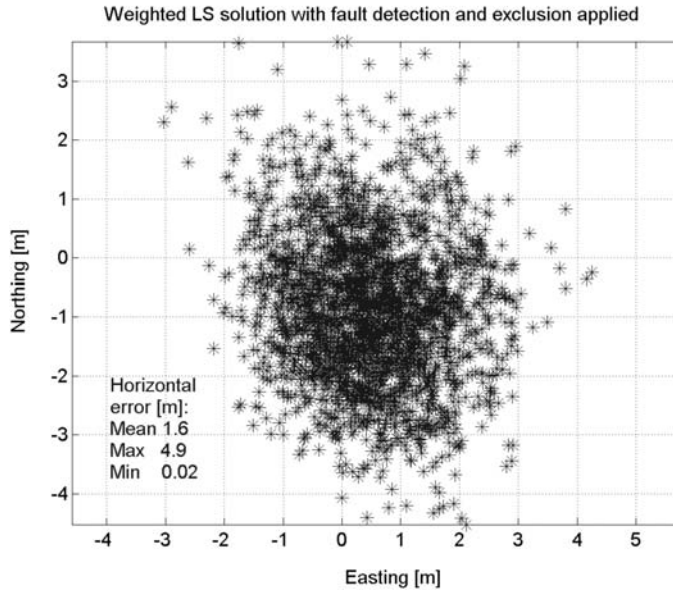


Figure 12.17 The WLS positioning result when fault detection and exclusion is applied.

tracking, data detection, and demodulation) as well as in the solution estimation (solution computation and quality monitoring) are required. The physical facts of obstructions cannot be denied, and at times, no amount of signal dwelling or error detection can ensure a navigation solution in the difficult signal environments often encountered indoors or in urban canyons. Therefore, backup systems for GNSS are necessary to be combined for an indoor and weak-environment application if the target is 100% location availability. This chapter outlines the different methods that should be considered in order to increase weak signal environment accuracy, reliability and availability.

12.8 Further Reading

There is a wide variety of significant references related to methods for indoor and weak signal-environment navigation. To explore this subject in greater depth, please refer to [4, 7, 18] for weak-signal and indoor processing methods, to [29] for various technologies of location-based services, to [22] for GPS pseudolites, to [43] for various traditional RAIM/FDE methodology, and to [33] for FDE in urban environments. Chapters 5–7 and 9 of [5] and Chapters 7, 8, and 15 of [65] also cover many of the discussed issues important for successful weak-signal navigation.

References

- [1] E112. Enhanced 112, 2005. URL <http://www.telematica.de/html/locus1.html>.
- [2] E911. Enhanced 911, Nov. 2004. URL <http://www.fcc.gov/911/enhanced/>.

- [3] GPS interface control document, ICD_GPS-200C, April 2000. URL <http://www.navcen.uscg.gov/pubs/gps/icd200/icd200cw1234.pdf>.
- [4] Dedes, G., and A. G. Dempster, "Indoor GPS Positioning, Challenges and Opportunities," *Proc. IEEE Vehicular Technology Conference*, Sept. 2005, pp. 412–415.
- [5] Kaplan, E. D., and C. J. Hegarty, eds., *Understanding GPS: Principles and Applications* (Second edition), Norwood, MA: Artech House, 2006.
- [6] Parkinson, B. W., and J. J. Spilker, eds., *Global Positioning System: Theory and Applications, Volume 1*, Washington, D.C.: AIAA Inc., 1996.
- [7] Ziedan, N. I., *GNSS Receivers for Weak Signals*, Norwood, MA: Artech House, 2006.
- [8] J. B.-Y. Tsui, *Fundamentals of Global Positioning System Receivers, A Software Approach* (Second Edition), New York, NY: John Wiley & Sons, 2005.
- [9] Watson, R., et al, "Investigating GPS Signals Indoors with Extreme High-Sensitivity Detection Techniques," *Navigation*, Institute of Navigation, Vol. 52, nr. 4, pp. 199–213.
- [10] Lombardi, M. A., "Fundamentals of Time and Frequency," Chapter 17 in *Mechatronic Systems, Sensors, and Actuators: Fundamentals and Modeling*, (ed. by R. H. Bishop, Boca Raton, FL: CRC Press, 2007. Available at <http://tf.nist.gov/timefreq/general/pdf/1498.pdf>.
- [11] Duffett-Smith, P. J., and A. R. Pratt, "Reconstruction of the Satellite Ephemeris from Time-Spaced Snippets," *Proc. ION GNSS 2007*, Fort Worth, TX, Sept. 2007, pp. 1867–1875.
- [12] Glennon, E. P., R. C. Bryant, and A. G. Dempster, "Delayed Parallel Interference Cancellation for GPS C/A Code Receivers," *Proc. IAIN/GNSS 2006*, Jeju, Korea, October 2006, pp. 261–266.
- [13] Madhani, P. H., et al, "Application of Successive Interference Cancellation to the GPS Pseudolite Near-Far Problem," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 39, nr. 2, Apr. 2003, pp. 481–488.
- [14] Lakhzouri, A., et al, "Interference and Indoor Channel Propagation Modeling Based on GPS Satellite Signal Measurements," *Proc. ION GPS 2005*, Long Beach, CA, Sept. 2005, pp. 896–901.
- [15] J. Ray, *Mitigation of GPS Code and Carrier Phase Multipath Effects Using a Multi-Antenna System*, Ph.D. thesis, Department of Geomatics Engineering, University of Calgary, Canada, 2000.
- [16] van Dierendonck, A. J., "Theory and Performance of Narrow Correlator Spacing in a GPS Receiver," *Navigation*, Institute of Navigation, Vol. 39, nr. 3, 1992.
- [17] Syrjärinne, J., *Studies of Modern Techniques for Personal Positioning*, Ph.D. thesis, Tampere University of Technology, Tampere, Finland, 2001.
- [18] van Diggelen, F., "Indoor GPS Theory and Implementation," *Proc. IEEE Position, Location, and Navigation Symposium*, Palm Springs, CA, Apr. 15–18, 2002, pp. 240–247.
- [19] Jokitalo, T., et al, "Performance Assessment of Almanac Navigation," *Proc. ION GNSS 2007*, Fort Worth, TX, Sept. 2007, pp. 1351–1358.
- [20] Lundgren, D., and F. van Diggelen, "Assistance When There's No Assistance, Long-Term Orbit Technology for Cell Phones, PDAs," *GPS World Magazine*, Oct. 2005.
- [21] Monnerat, M., "AGNSS Standardization: The Path to Success in Location-Based Services," *Inside GNSS Magazine*, July/August 2008, pp. 22–33.
- [22] Cobb, H. S., *GPS Pseudolites: Theory, Design, and Applications*, Ph.D. thesis, Stanford University, CA, 1997.
- [23] Söderholm, S., and T. Jokitalo, "Synchronized Pseudolites—The Key to Indoor Navigation," *Proc. ION GNSS 2002*, Sept. 2002, pp. 226–230.
- [24] Mezentsev, O., *Sensor Aiding of HSGPS Pedestrian Navigation*, Ph.D. thesis, Department of Geomatics Engineering, University of Calgary, Calgary, Canada, 2005.
- [25] Collin, J., *Investigations of Self-Contained Sensors for Personal Navigation*, Ph.D. thesis, Tampere University of Technology, Finland, 2006.
- [26] Savage, P. G., *Strapdown Analytics*, Maple Plain, MN: Strapdown Associates, Inc., 2000.

- [27] Grewal, M. S., L. R. Weill, and A. P. Andrews, *Global Positioning Systems, Inertial Navigation, and Integration*, New York: John Wiley & Sons, Inc., 2001.
- [28] Kolodziej, K. W., and J. Hjelm, *Local Positioning Systems: LBS Applications and Services*, Boca Raton, FL: CRC Press, Taylor & Francis Group, 2006.
- [29] Küpper, A., *Location-Based Services: Fundamentals and Operation*, New York: John Wiley and Sons, 2005.
- [30] Roos, T., et al, "A Probabilistic Approach to WLAN User Location Estimation," *International Journal of Wireless Information Networks*, Vol. 9, No. 3, 2002, pp. 155–163.
- [31] Sirola, N., *Mathematical Methods for Personal Positioning and Navigation*, Ph.D. thesis, Tampere University of Technology, Finland, 2007.
- [32] Rabinowitz, M., and J. J. Spilker, "Augmenting GPS with Television Signals for Reliable Indoor Positioning," *Navigation*, Institute of Navigation, Vol. 51, No. 4, Winter 2004, pp. 269–282.
- [33] Kuusniemi, H., et al, "User-Level Reliability Monitoring in Urban Personal Satellite-Navigation," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 43, No. 4, October 2007, pp. 1305–1318.
- [34] Peterson, B., R. Hartnett, and G. Ottman, "GPS Receiver Structures for the Urban Canyon," *Proc. ION GPS 1995*, Sept. 1995, pp. 1323–1332.
- [35] Quddus, M. A., W. Y. Ochieng, and R. B. Noland, "Current Map-Matching Algorithms for Transport Applications: State-of-the Art and Future Research Directions," *Transportation Research*, Part C, Emerging Technologies, 2007, Vol. 15, No. 5, pp 312–328.
- [36] Kuusniemi, H., *User-Level Reliability and Quality Monitoring in Satellite-Based Personal Navigation*, Ph.D. thesis, Tampere University of Technology, Finland, 2005.
- [37] Brown, R. G., and P. Y. C Hwang, *Introduction to Random Signals and Applied Kalman Filtering* (Third Edition), New York, NY: John Wiley & Sons, Inc., 1997.
- [38] Brown, R. G., "A Baseline RAIM Scheme and a Note on the Equivalence of Three RAIM Methods," *Proc. ION NTM 1992*, San Diego, CA, January 1992, pp. 127–138.
- [39] Brown, R. G., "Receiver Autonomous Integrity Monitoring," Chapter 5 in *Global Positioning System: Theory and Applications Volume 2*, B. W. Parkinson and J. J. Spilker, eds., Washington, D.C.: AIAA Inc., 1996, pp. 143–165.
- [40] Ober, P. B., "Integrity Prediction and Monitoring of Navigation Systems," Ph.D. dissertation, Delft Univ. Tech., The Netherlands, 2003.
- [41] Van Graas, F., and J. L. Farrell, "Baseline Fault Detection and Exclusion Algorithm," *Proc. ION AM 1993*, Cambridge, MA, 1993, pp. 413–420.
- [42] Kelly, R. J., "The Linear Model, RNP, and the Near-Optimum Fault Detection and Exclusion Algorithm," paper published in *NAVIGATION, Journal of The Institute of Navigation*, Vol. 5, 1998, pp. 227–259.
- [43] "Red Book of RAIM," Special ION Publication on RAIM, Global Positioning System, paper published in *NAVIGATION*, Volume V, Institute of Navigation, Alexandria, VA, 1998.
- [44] Misra, P., and P. Enge, *Global Positioning System, Signals, Measurements, and Performance*, Lincoln, MA: Ganga-Jamuna Press, 2001.
- [45] G. Strang and K. Borre, *Linear Algebra, Geodesy, and GPS*, Wellesley, MA: Wellesley-Cambridge Press, 1997.
- [46] Borre, K., et al, *A Software-Defined GPS and Galileo Receiver, A Single-Frequency Approach*, Boston: Birkhäuser, 2007.
- [47] Hawkins, D. M., *Identification of Outliers*, London/New York: Chapman & Hall, 1980.
- [48] Kay, S. M., *Fundamentals of Statistical Signal Processing: Estimation Theory*, Englewood Cliffs, N.J.: Prentice Hall International, Inc., 1993.
- [49] Hartinger, H., and F. Brunner, "Variances of GPS Phase Observations: the SIGMA- ϵ Model," *GPS Solutions*, Vol. 2, No. 4, Apr. 1999, pp. 35–43.

- [50] Teunissen, P. J. G., "Quality Control and GPS," in *GPS for Geodesy* (Second Edition), P. J. G. Teunissen and A. Kleusberg, eds., New York, NY: Springer, 1998, pp. 271–318.
- [51] Kuang, S., *Geodetic Network Analysis and Optimal Design*, Chelsea, MI: Ann Arbor Press, 1996.
- [52] Chaffrin, B., "Reliability Measures for Correlated Observations," *Journal Surv. Eng.*, Vol. 123, August 1997, pp. 126–137.
- [53] Leick, A., *GPS Satellite Surveying* (Third Edition), New York: John Wiley & Sons, Inc., 2004, Ch. 4, pp. 92–169.
- [54] Koch, K.-R., *Parameter Estimation and Hypothesis Testing in Linear Models* (Second Edition), Berlin, Heidelberg: Springer-Verlag, 1999.
- [55] Baarda, W., *A Testing Procedure for Use in Geodetic Networks*, Netherlands Geodetic Commission, Publication on Geodesy, New Series 2, 5, Delft, Netherlands, 1968.
- [56] Rousseeuw, P. L., and A. M. Leroy, *Robust Regression and Outlier Detection*, New York: John Wiley & Sons, Inc., 1987.
- [57] Caspary, W. F., *Concepts of Network and Deformation Analysis*, School of Surveying, University New South Wales, Australia, Monograph, 11, 1988.
- [58] Petovello, M. G., *Real-Time Integration of a Tactical-Grade IMU and GPS for High-Accuracy Positioning and Navigation*, Ph.D. dissertation, University of Calgary, Canada, 2003.
- [59] Lu, G., *Quality Control for Differential Kinematic GPS Positioning*. M.Sc. thesis, University of Calgary, Canada, 1991.
- [60] Kuusniemi, H., G. Lachapelle, and J. Takala, "Position and Velocity Reliability Testing in Degraded GPS Signal Environments," *GPS Solutions*, Vol. 8, No. 4, Dec. 2004, pp. 226–237.
- [61] Wieser, A., *Robust and Fuzzy Techniques for Parameter Estimation and Quality Assessment in GPS*, Ph.D. dissertation, Technical University Graz, Austria, 2001.
- [62] Angus, J. E., "RAIM with Multiple Faults," *Navigation*, Institute of Navigation, Vol. 53, No.4, Winter 2006, pp. 249–257.
- [63] Grewal, M. S., and A. P. Andrews: *Kalman Filtering*, New York: John Wiley and Sons Ltd., 2001.
- [64] Wieser, A., M. G. Petovello, and G. Lachapelle, "Failure Scenarios to Be Considered with Kinematic High Precision Relative GNSS Positioning," *Proc. ION GNSS 2004*, Long Beach, CA, Sept. 2004, pp. 1448–1459.
- [65] Groves, P. D., *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*, Norwood, MA: Artech House, 2008.

Space Applications

E. Glenn Lightsey

13.1 Introduction

Although it is principally designed for terrestrial users, GNSS technology has been broadly adopted by the spacecraft industry. GNSS receivers provide a compelling list of potential capabilities, including autonomous real-time navigation, centimeter-level postprocessed orbit determination, atomic clock timing accuracy, relative navigation between orbiting vehicles, and attitude determination. All this comes in a small electronic device that requires just a few watts of power to operate. The satellite community quickly embraced these potential benefits, and spacecraft have employed GNSS services since the first GPS satellites became operational. Landsat-4 was among the first satellites to report results from a designed-for-space GPS receiver in 1982 [1]. Spacecraft reliance on these systems has grown steadily over time. The once expensive, bulky, and somewhat limited receivers have become more affordable (in some cases), smaller, and more capable, allowing GNSS receivers to be employed in many different space applications with varied requirements and budgets.

It is not practical to list all past and present space missions using GNSS in a single chapter, nor is it possible to anticipate every possible future application. Instead, this chapter summarizes the distinguishing characteristics of the space environment and discusses how they affect the design and performance of space GNSS receivers in general. The chapter then briefly reviews the most common and the most promising space applications to indicate the opportunities that lay ahead for space GNSS users in the next decade. It is clear that GNSS receivers will continue to become more prevalent on spacecraft of all types and that more advanced features and designs will continue to emerge over the next decade.

13.2 Operational Considerations

A common misperception is that all GNSS receivers are functionally the same and that a receiver that is well-designed for terrestrial applications will also work well in space. Unfortunately this assumption is false, and most terrestrial receivers would not work at all in space for a variety of reasons related to the different operating environment. Several of these differences are briefly discussed.

13.2.1 Spacecraft Velocity

A GNSS receiver installed on an orbiting satellite moves with a velocity of several kilometers per second. The greater relative velocity between the GNSS receiver and the transmitting GNSS satellite induces a correspondingly greater Doppler shift in the signal that the GNSS receiver must track, and requires the receiver to search over a large range of frequencies to acquire the GNSS signal. For a typical GPS receiver operating on a LEO satellite, the range of possible incoming Doppler shifts may be extended from approximately 5–10 kHz for terrestrial users to more than ± 30 kHz in space. For example, Figure 13.1 shows one hour of Doppler shift data measured by a GPS receiver on the RADCAL satellite, which was launched into an approximately 800-km polar orbit [2]. This larger Doppler shift range makes the initial acquisition problem more difficult without assistance and lengthens the time to first position fix.

To accelerate the acquisition of GNSS signals in space, many receivers employ special algorithms or added hardware. For instance, given a set of orbit elements and current time, the receiver's coarse position and velocity may be predicted. This is similar to the position initialization of a terrestrial receiver. The satellite search list and nominal Doppler offset frequency can be adjusted based on this information. Some new receivers are also employing software-defined radio techniques to perform signal acquisition within seconds, even without prior knowledge of the receiver's position [3].

13.2.2 Orbit Geometry

The orbit of the receiver's satellite strongly influences the time history of visible GNSS satellites. GNSS satellites are placed into orbits that are designed to maximize

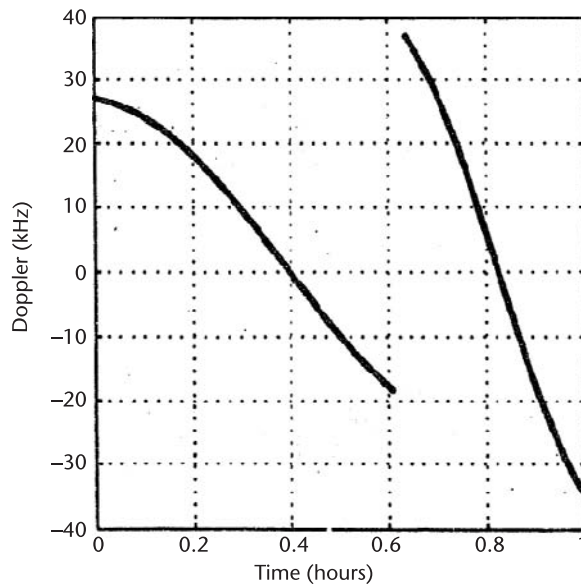


Figure 13.1 Doppler shift of GPS signal on the RADCAL satellite. The discontinuity at 0.6 hours is due to a switch in tracked signals [2].

their visibility to terrestrial users. These transmitting satellites have orbits with periods of many hours. For example, GPS satellites are visible to a ground user with an unobstructed sky view for at least eight hours, or 480 minutes. This long-term visibility affects the nature of the receiver satellite selection algorithm and associated signal tracking dynamics. In contrast, a satellite placed into LEO with a GNSS receiver may have an orbital period of less than 100 minutes. In this case, from the receiver's perspective, the visible satellites may change from one hemisphere to another in less than 50 minutes, which is roughly an order of magnitude faster than for a ground-based receiver. This necessitates a change to the satellite selection logic and tracking loop design for the receiver to quickly acquire rising signals and to assure good geometric diversity of received signals.

Higher altitude orbits pose other challenges for orbit geometry. Although the periods of these orbits are longer, traditional methods for determining which satellites are visible to the receiver by computing the direction to the local vertical may not be successful if the altitude of the receiving satellite is near or above the altitude of the transmitting GNSS satellites. In these cases, the GNSS receiver may be able to acquire the signals from satellites transmitting across the Earth, if the main beam of the transmitting antenna pattern is wide enough to permit reception (see Figure 13.2). Alternatively, it may be possible to track signals from the much weaker sidelobes of the transmitting satellites.

The special case of a HEO (as shown in Figure 13.2) combines the properties of LEO signal tracking at perigee with high-altitude satellite selection at apogee. These challenging orbits (consider for example, NASA's proposed MMS mission [5]) require satellite selection strategies that smoothly scale from one extreme to the other based on the receiver position.

Regardless of which situation applies to an individual project, the GNSS receiver's satellite selection logic and its resulting on-orbit performance are strongly influenced by the receiver's orbit. The ideal receiver must have modifiable tracking loop assignments to be able to handle these varying cases. Software receivers, which have many customizable features, may be well-suited for these considerations.

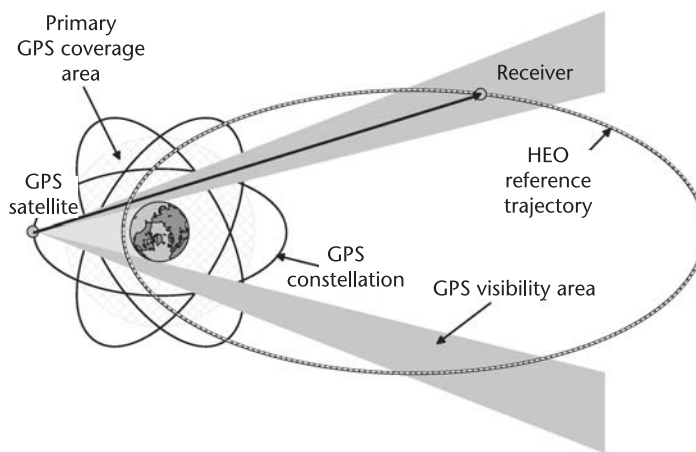


Figure 13.2 Visibility of GPS satellites to a receiver in a high-altitude elliptical orbit (HEO) [4].

Programmable hardware (as employed, for example, in [6]) can also enable advanced receiver architectures and signal processing capabilities in space.

13.2.3 Antenna Direction

Most terrestrial receivers are designed to be used over a relatively small range of non-azimuthal rotation. Aircraft, which have more rotational freedom than most ground-based vehicles, rarely bank more than 45 degrees from vertical during normal flight conditions. For terrestrial users (including aircraft), it is possible to install a GNSS receiver antenna in an upward facing direction and assume it will be reasonably well-pointed in that direction at all times.

However, a GNSS receiver installed on a spacecraft must be prepared to receive signals from all directions in the vehicle's reference frame unless the platform's attitude is stabilized by some separate means. This so-called full sky reception can be accomplished by installing multiple antennas pointed in different directions, or by modifying the antenna's gain pattern. Alternately, the designer may choose to consciously restrict the receiver's potential field of view in the vehicle reference frame. In that case, the satellite's design must be able to accept the possibility of GNSS signal outages due to the attitude motions of the vehicle.

Some spacecraft are spin-stabilized during a portion of their mission life. If a GNSS receiver is needed during this mission phase, some aspect of the receiver's design must be altered to account for the spinning motion. This can be achieved through hardware or software design accommodations. In a hardware solution, the antenna may be pointed along the spin axis or the antenna gain pattern may be modified, for example by using a hoop or wrap-around array of antenna elements. If a software solution is desired, satellite selection logic and tracking loops may be adjusted to account for the spinning motion. Custom satellite selection algorithms have been developed for spinning and other types of orbiting platforms [7, 8].

13.2.4 Size and Power

Although not unique to space applications, size and power are at a premium on all space vehicles. Depending on the satellite, there may be a volume requirement for the GNSS receiver to fit within a mechanical envelope, which may be as small as a few cubic centimeters. There is competition from many other subsystems for available area on the surface of the vehicle, so a small GNSS antenna footprint is also desirable. In order to obtain a working vehicle design, it may be necessary to limit the number of GNSS antennas and/or restrict their fields of view. This will result in increased multipath noise and possible signal blockages that must be accommodated operationally.

In some cases, the power of the receiver may be a limitation. Although GNSS receivers are not inherently high in power consumption, even a few watts of orbit average power may be too high for some of the smallest and lowest power satellites. However, most satellites have batteries that enable short periods of greater power consumption, allowing receiver operation to be duty-cycled over an orbit if necessary. Some innovative low-power receiver solutions have been proposed, such as receiver bit grabbers that record signals only for portions of an orbit [9].

13.2.5 Multipath

Although a satellite operating in space does not generally receive reflections from other objects, multipath can be self-induced from the nearby surfaces on the vehicle. Because the surfaces on a satellite are often complex in shape, and the designer has limited ability to place the antenna to obtain the best performance, the amount of multipath affecting a GNSS measurement in space can be severe. This is especially true on large complex structures with many reflective objects and surfaces, such as the International Space Station (ISS) [10]. Also, when two or more vehicles are performing a coordinated proximity maneuver such as formation flying or docking, it is possible that one of the vehicles will become a reflection source or block the field of view of the other vehicle's GNSS antennas (Figure 13.3). Thus unfortunately, the performance of these systems may be the most vulnerable at exactly the time that the greatest level of accuracy and reliability is needed.

Depending on the need, it may be possible to partially mitigate these considerations by optimizing the antenna placement and by using practices such as choke rings to restrict reflections. However, these options come at the expense of placement of other subsystems and may not always be available. For example, choke rings require more of the vehicle's limited surface area to work properly.

13.2.6 Signal Strength

For LEO applications (up to approximately 3,000 km altitude), the direct LOS received signal power is close to the normal levels of variability encountered by a terrestrial user.

Therefore it is not necessary to adjust the antenna gain pattern for normal use. This region has been described as the terrestrial service volume [11]. However, for some applications, such as occultation measurements (see Chapter 15), special

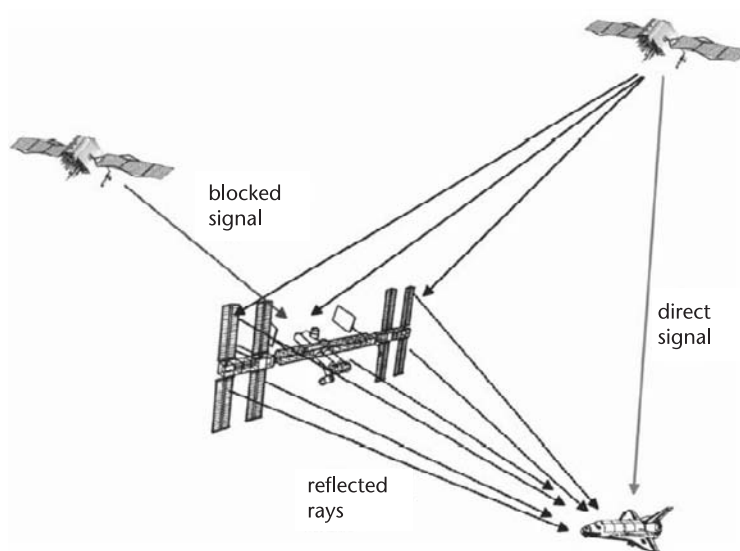


Figure 13.3 Example proximity signal blockage and multipath scenario in space [10].

antennas may be employed to track the signal through the limb of the Earth’s atmosphere. These signals are attenuated compared to direct signals, and the antennas that track them are designed to have greater sensitivity near the horizon [12].

For higher altitude applications near and above the GNSS altitude (known as the space service volume), the transmitting antenna gain pattern and free-space transmission loss must also be considered. In most cases, these signals are weaker due to the greater distances and suboptimal transmitting and receiving antenna gains. By using a directional antenna with higher receiving gain, it is still possible to track these signals. Tracking of GPS signals, for example, has been successfully demonstrated on several high-orbit experiments [13, 14]. Figure 13.4, showing flight data taken from the AO-40 satellite, illustrates that some of the GPS block IIR satellite sidelobe signals are transmitted at higher power levels than specified.

13.2.7 Environment

The space environment is one of the most extreme, and this also influences the design of a GNSS receiver, which must operate on a satellite. The main issues are surviving launch and possible re-entry vibration loads, thermal cycling over a wide temperature range, and exposure to hard vacuum and radiation. These considerations are common for all space hardware and electronics.

Depending upon the mission requirements, aviation-grade commercial off-the-shelf electronics such as those found in terrestrial receivers may be sufficient as long as proper precautions are taken. For instance, plastics and other materials that outgas in a vacuum are typically conformally coated. Surface-mounted parts may be potted and staked to withstand vibration loads. Parts may be selectively replaced for thermal and radiation survivability. Radiation testing has shown that some GNSS receivers built using standard aviation grade parts can withstand more than 9 kilorads total ionizing dose without fault [15, 16]. This is sufficiently hardened for multiyear LEO missions but requirements of 100 kilorads for higher orbit altitudes necessitate different parts. These considerations have to be made on a case-by-case basis and drive the cost and selection of GNSS receiver hardware for various missions. For example, a GNSS receiver that must provide a guaranteed

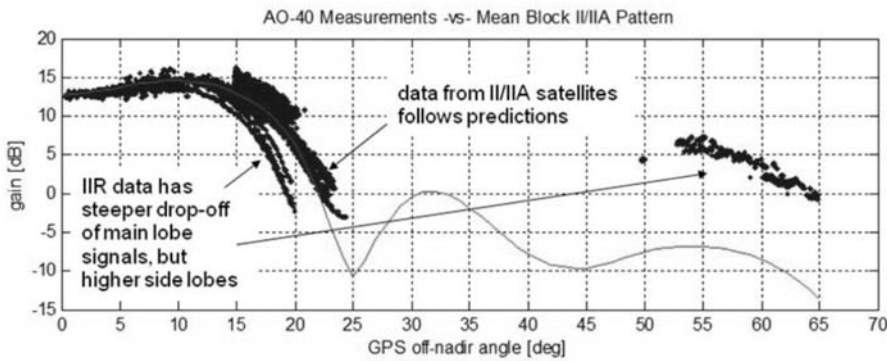


Figure 13.4 Received signal power measurements taken at high altitudes on the AO-40 satellite [13].

life of more than five years operating in the Earth's radiation belts will have totally different design and cost than an experimental receiver with no minimum guaranteed lifetime designed for operation in LEO.

13.3 Applications

Considering the variety of space missions and their differing requirements, there must naturally also be a variety of GNSS receivers to fill the needs of these different missions. The following sections briefly discuss several different mission functions that GNSS receivers have already fulfilled or are expected to perform within the next decade.

13.3.1 Precise Orbit Determination

Orbit determination using radio transmission has been employed since the first man-made satellites. Continuing improvements in measurements and processing methods have created a technique that is known today as Precise Orbit Determination (POD). Dual-frequency GNSS receivers that are designed for POD have flown on many satellites when accurate knowledge of the vehicle orbit was required to meet mission objectives. Processing methods have been optimized during missions like TOPEX/Poseidon, and current levels of achievable orbit accuracy are between 0.5 cm and 2 cm rms radial error. Performance depends on the measurement environment, the processing techniques used, and the mission application. For example, the Jason mission has demonstrated GPS-reduced dynamic model solutions with radial orbit accuracy better than 1 cm [17], and the TerraSAR-X satellite has recently obtained approximately 1–2-cm rms radial orbit accuracy and better than 5-cm three-dimensional rms position [18]. Independent verification of navigation performance beyond these levels is challenging with present technology.

While accuracy remains the defining characteristic of POD methods, there is growing emphasis on shortening the time latency required to achieve precise solutions. Many end users benefit from having orbit solution products that are rapidly available. Some research groups are focusing on providing predicted trajectories as soon as possible, with a goal of achieving near real-time navigation solutions with POD levels of accuracy.

13.3.2 Real-Time Navigation

GNSS services also function as a real-time sensor on the vehicle for navigation, guidance, and control, and in this case it is not usually possible to employ the ground station networks that are used with POD products. For LEOs, kinematic point position solutions are very similar in accuracy to those found in terrestrial applications (~ 10 -m position and ~ 0.1 -m/s velocity), subject to signal availability and dilution of precision considerations. However, it is possible to obtain better performance by adopting real-time filtering methods that incorporate a dynamic model of the orbit. For example, one study has shown that half-meter rms real-time navigation accuracy is achievable using a dynamic filter with dual frequency

GPS measurements (see Figure 13.5) [19]. Performance varies with orbit conditions, as expected. The accuracy of the filtered solution is dominated by the position error of the transmitting GNSS satellites. If the GNSS satellite positions are known more accurately than what is available from the broadcast ephemerides, even better performance is possible. Although there are not many low-cost dual-frequency receivers today, these types of receivers are expected to become more widely available in the future with software receivers, GNSS signal modernization, and with the addition of new GNSS constellations.

13.3.3 Formation Flying and Proximity Operations

The real-time navigation capability of GNSS provides for many space formation flying and proximity operations possibilities. In the simplest case, known as a swarm, each vehicle broadcasts measurements from which its position may be determined. Because the measurements that the satellites share may have correlated errors, the relative position between receiving satellites may be determined more accurately in some cases than the absolute position. This may be beneficial for guidance or postprocessing. In this case, however, the vehicle positions are not controlled relative to each other.

If controlled flight between vehicles is intended, at least one of the vehicles must possess and act upon the relative dynamic state information. The usual method of data exchange is that one vehicle receives measurements that are transmitted from the other and computes a relative solution, although this is not strictly required as long as the relative navigation information is conveyed by some means. Once

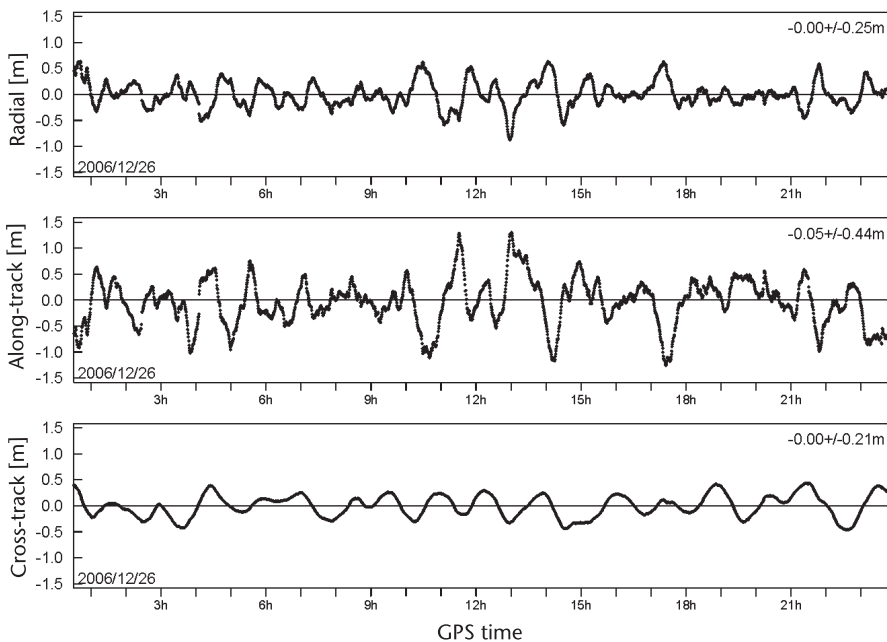


Figure 13.5 Filtered GNSS navigation accuracy on the MetOp satellite (820 km altitude, 98.7 degree inclination) using dual frequency measurements with the broadcast ephemeris.

the relative state is known, formation flying and cooperative operations are possible. Some examples of possible missions are satellite inspection, rendezvous and docking, on-orbit servicing, and coordinated flight for scientific and other applications.

A number of formation flying and proximity missions have been performed that were enabled by GNSS receivers. One interesting example of a combined system is GRACE. In this system of two satellites, the GNSS receiver performs several functions. The navigation solution is used to provide guidance and targeting between the two vehicles to maintain a nominal separation distance of approximately 220 km. The formation control of GRACE is a ground-based process with time scales of several months between maneuvers. The real-time use of GPS is restricted to enabling onboard timing as well as for providing the inertial orientation of the orbital frame (i.e., the nadir and forward direction) for attitude control. POD is also obtained from the GNSS measurements in postprocessing to provide centimeter-level registration of each satellite's absolute position. A K/Ka microwave ranging system provides a more sensitive micron-level distance measurement between the two vehicles. The extremely sensitive relative navigation measurements have been used to map the Earth's geoid to an unprecedented accuracy [20].

Because the coordinated flight of multiple vehicles is inherently more challenging than the guidance of a single vehicle, this capability is still emerging. The autonomous coordinated flight of spacecraft in proximity to each other is a highly complex mission. Proximity vehicle blockage of GNSS signals and attitude motions of each vehicle must be taken into account. DART is an example of one mission that experienced an unanticipated interaction between a GNSS receiver and the system's guidance and control system. As a result, the maneuvering vehicle exhausted its propellant before it was able to accomplish its proximity objectives. An unforeseen interaction between the sensors and the controller caused the two vehicles to collide during a maneuver [21]. This experience demonstrates the difficulty of autonomous operations and the challenge of fully characterizing the behavior of a complex system prior to flight. Other missions with more basic objectives have demonstrated some level of coarse coordinated flight using low-cost GNSS receivers (for example, the Tsinghua-1/SNAP-1 experiment [22]), and progress in this area is expected to continue.

13.3.4 Remote Sensing

In addition to providing navigation services, the measurements collected from GNSS receivers may be used directly for remote sensing. Two examples that have been demonstrated are atmospheric occultation and bistatic radar. Both of these applications are covered in more detail in Chapters 15 and 16, respectively.

By measuring the refraction angle and delay of radio signals traveling through the limb of the Earth's atmosphere, it is possible to infer properties about the state of the atmosphere, such as: temperature, density, water content, and composition [23]. Because orbiting platforms have the ability to sample large sections of the atmosphere from many different locations over time, it is natural that GNSS receivers would be employed for this purpose. These receivers have been specially adapted for occultation measurements. For example, the antennas are often horizon-looking and the receiver tracking logic is adjusted to select lower elevation

measurements. Some examples of atmospheric occultation experiments are: GPSMET, CHAMP, SAC-C, and COSMIC/Formosat-3. COSMIC/Formosat-3 is a constellation of six satellites each orbiting between 400 and 800 km altitude and 72 degrees inclination [24].

Reflected GNSS signals can also be used as a bistatic radar measurement. The time delay of the reflected signal can determine the height above the reflection source, and the spectral properties of the reflected signal can provide information about the characteristics of the reflection surface. For example, a strong coherent return signal indicates a specular reflection from a smooth surface, whereas a weak noncoherent return signal indicates a diffuse reflection from a rough surface. A suitably equipped GNSS receiver with a high gain antenna may use bistatic radar measurements to study the Earth. Some of the phenomena that can be studied are sea-state roughness, which is a function of surface winds, and ground surface properties including possibly soil moisture. The GPS bistatic radar concept has been shown to be feasible on LEO spacecraft for sensing ocean, land, and ice, although additional validation is still needed [23].

13.3.5 Attitude Determination

Although GNSS are primarily used for navigation and time distribution, algorithms have also been developed that allow its use for attitude determination. These algorithms generally fall into two categories: multiantenna platforms that use carrier-phase interferometry between two or more antennas, and single-antenna systems that use variations in the measured received signal power. In both cases, the object of the measurement is to determine the platform's orientation relative to the transmitting satellite's LOS direction. Since the transmitting satellite's direction is known in an external reference frame (assuming that the position of the receiving platform is known), and the direction of the antenna boresight is also known in the body fixed reference frame, each measurement may be correlated with the attitude of the platform's body fixed reference frame relative to the external reference frame.

For most spacecraft missions to date, GNSS attitude determination has been a relatively coarse sensor providing degree level accuracy, depending on the application. The main limitations in accuracy are the measurement errors primarily due to multipath, and for the case of multiantenna interferometry systems, the size of the antenna array, which is rarely more than 1m. The carrier-phase interferometry method contains a cycle ambiguity that becomes more difficult to solve as the baseline length between antennas increases.

Despite the challenges, GNSS attitude determination has been successfully employed on several spacecraft. For example, the ISS has used a GPS receiver to provide attitude solutions since 2002 with a 1.5m-by-3m four-antenna array. After recoding portions of the receiver software to improve performance, the custom-designed GPS system has provided attitude solutions that are estimated to be accurate within 4 degrees rms in the severe multipath environment of ISS [25]. In 1996, the REX-II satellite successfully closed its attitude control loop using GPS attitude solutions [26].

For a single-antenna GNSS attitude sensor, the important metrics are the accuracy of the SNR measurement (which is perturbed by multipath and is also highly quantized in some GNSS receivers) and the knowledge of the receiver antenna gain pattern. Although this approach is generally not as accurate as the interferometry approach, it requires less hardware to operate, and still provides coarse attitude knowledge, which may be acceptable for some situations. For example, a single-antenna GPS attitude sensor was used on FedSat to determine the vehicle's attitude within 15 degrees rms (see Figure 13.6) [27]. Performance may be improved by using a well-known or calibrated antenna gain pattern and employing a receiver that is designed to make accurate signal strength measurements. It may also be desirable or necessary to combine the GNSS measurements with another independent sensor, such as a magnetometer or Sun sensor, to provide additional solution robustness and three-axis attitude determination.

13.3.6 High-Altitude GNSS

As previously discussed, GNSS receivers operating in and above the region designated as the space service volume (orbit altitude above 3,000 km) must function in a markedly different environment than those operating in the terrestrial service volume (altitude less than 3,000 km). Compared to GNSS receivers designed for LEO, these receivers must have modified tracking and satellite selection logic, different antenna design and placement, and radiation-hardened electronics. Because of the greater distance from Earth, the number of visible satellites is less and the geometric diversity of these measurements (as indicated through dilution of precision) is relatively poor. Nonetheless, it is possible to track signals in high orbits [13, 14]. By using dynamic filtering and postprocessing along with possible additional sensors, navigation solutions ranging in accuracy from tens of meters to meter-level are possible. Tracking of Earth-orbiting GNSS satellites has even been proposed for lunar missions, although this remains a topic of current research [28].

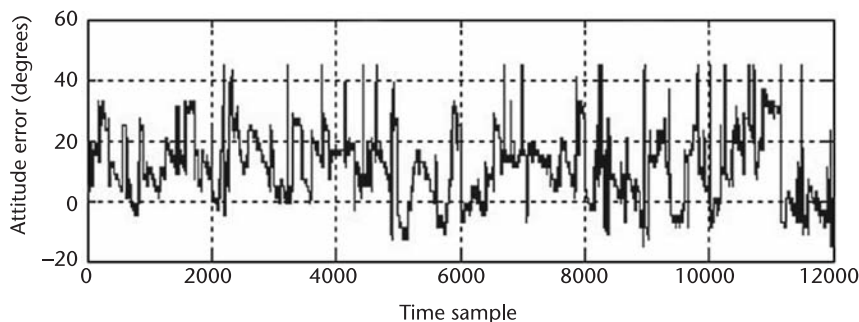


Figure 13.6 Attitude solutions computed from the FedSat satellite (803-km altitude, 98.7-degree inclination) using a single-antenna GPS receiver method [27]. Figure reproduced with permission and special thanks due to Charles Wang, Rodney Walker and Miles Moody at Queensland University of Technology.

13.3.7 Launch, Entry, and Landing

A special case occurs for GNSS receivers that must operate in high-acceleration environments that are encountered during launch, ascent, and atmospheric entry. Depending on the motion encountered, this may require designing the tracking loops to provide for a higher dynamic range than in standard space receivers [29]. Antenna mounting and signal visibility are additional considerations. It is common practice to combine these receivers with inertial navigation systems to provide higher rate measurements and added tracking robustness. For re-entry, a plasma induced radio blackout period will occur for several minutes in the upper atmosphere. It is usually critical for these systems to reacquire the GNSS signals as soon as possible to minimize the duration of the GNSS outage.

GNSS receivers are an important element of future launch range design [30]. Implementation in this area is presently ongoing. For example, the U.S. government is planning to incorporate GPS metric tracking for all launches from its eastern and western launch ranges by 2011 [31]. This policy will encourage all launch vehicles to be equipped with GNSS receivers, among other changes. Integrating space-based tracking with existing range practices such as ground-based radar will provide launch ranges with new measurements and is envisioned to reduce operating costs.

13.4 GNSS Modernization

Many changes are occurring in the GNSS industry presently and over the next decade that may be collectively referred to as GNSS modernization. These topics may be broadly classified into two categories: GPS legacy signal modernization and the introduction of new independently operated GNSS services. In both cases, the modifications should preserve backward compatibility with established users while providing better levels of navigation accuracy and more options for users equipped with new receivers.

GPS signal modernization is being phased in over the next decade and includes the addition of two new civilian codes known as L2C and L5, which will allow users to perform multifrequency ionosphere-free measurement combinations without requiring codeless P/Y signal tracking techniques. The difference will be most significant for occultation measurements, which suffer from semi-codeless tracking losses in signal strength. First on-orbit flight demonstrations of L2C tracking have been reported on SAC-C and COSMIC/Formosat-3 [32].

There will also be a new military code, known as M-code, which presently has unknown availability for space applications. However, ground stations that use the M-code to track the GPS satellites should be able to provide more accurate ephemeris solutions, which may reduce overall user equivalent range error. The transmitted power level of the GPS signal will also be increased, which will improve pseudorange measurement accuracy (see Figure 13.7 [33]).

New GNSS networks that are currently in development include the existing Russian GLONASS, the European Galileo positioning system, and the Japanese QZSS system. Other systems will possibly be introduced over time [34]. While designed for compatibility with existing positioning systems, the new networks

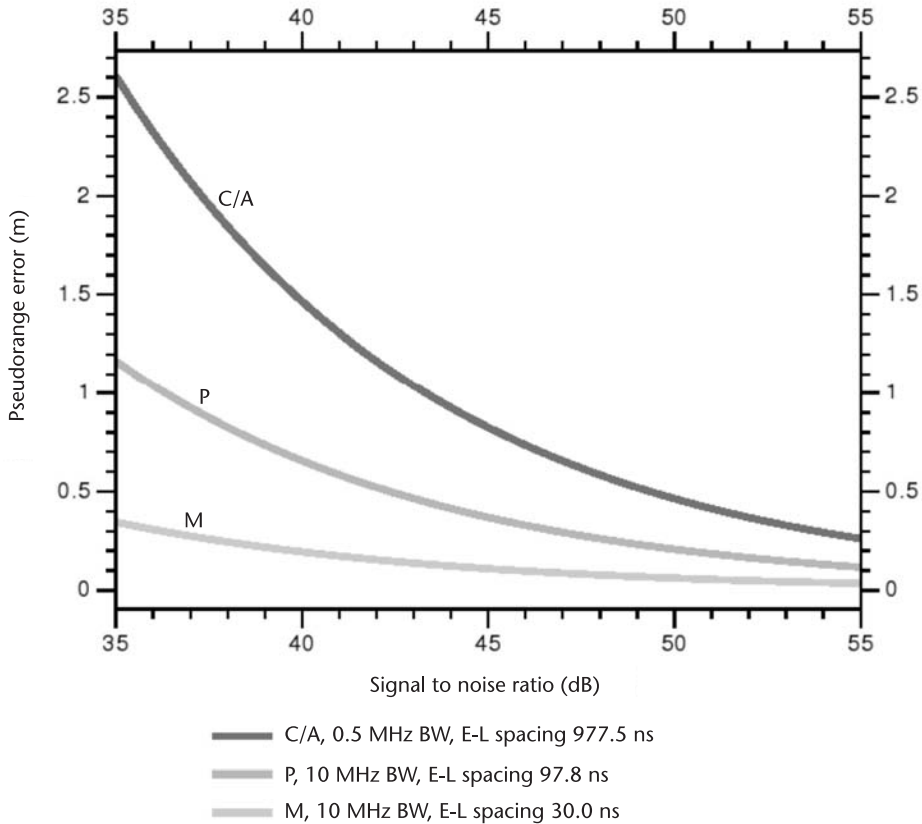


Figure 13.7 Pseudorange error as a function of signal to noise ratio for C/A, P, and M codes [33].

will provide more signals, frequencies, and services to GNSS receivers that are able to track them. As these systems become fully operational, it is expected that new hybrid GNSS receiver designs will be created that will be able to track signals from multiple GNSS services. Software receivers may be an enabling technology in this regard because they have the potential to be reconfigurable in flight.

Although many of the improvements that are anticipated have not yet been fully implemented and are therefore subject to budget and schedule constraints, the trend is clearly towards more GNSS signals and more GNSS services. This is beneficial not only because a greater diversity of measurements and more accurate measurements will be available, but also because global reliance on any single navigation carrier to support the space user community will be reduced.

13.5 Example: Processing Raw Measurements from the GRACE Satellite

Included on the DVD are raw dual-frequency pseudorange measurements collected from the NASA/JPL Blackjack receiver on the GRACE satellite. The main distribution Web site, including raw data and documentation for GRACE can be viewed at: <http://podaac.jpl.nasa.gov/grace/index.html>.

The data included on the DVD consisted of level 1B data collected on September 30, 2008. The zipped raw binary file and processing tools can be found in the “GRACE_Data_and_Tools” directory for this chapter. Included is a tool from the above Web site for generating a RINEX 2.20 file, including raw pseudorange measurements from all the satellites tracked by the Blackjack receiver. The entire set of data files extracted from the raw data are included in “Processing_GRACE_Data” directory. The RINEX file contains part of the information needed to compute a least-squares position estimate using the raw measurements, notably the time, satellite ID, and pseudorange measurements. The additional information that is needed to perform a position estimation includes: the transmitting satellite positions and clock bias information at the time of the measurements.

These data were processed by the receiver, but unfortunately are not included in the raw data packets. However, this data can be retrieved after the fact using archived ephemeris from the International GNSS Service (<http://igsceb.jpl.nasa.gov/>) and processed with the fitting tools written by Steve Hilla at NOAA (C++ version included in the software receiver from Chapter 5). A new file is then created by combining the GRACE RINEX file with satellite position and clock bias information. This new file can be found in the “Augmented_GRACE_GPS_Data” directory. This file was generated using a SWIG-generated Python wrapper to a slightly modified version of the satellite interpolation routines contained in the original C++ file. (For those interested, the code has been included in the “SP3_SWIG” subdirectory.) The resulting file can be processed using the MATLAB/Octave positioning script in the same directory, which is a slightly modified version of the estimation script included in Chapter 3.

When the GRACE_PVT.m script is run in MATLAB or Octave, it will estimate and plot orbital locations of the GRACE satellite over one orbit of the raw data set, illustrated in the Figure 13.8. A couple of comments are in order about these results. Note that a position fix drop out occurs for several seconds about 900 seconds into the data set--this drop out physically occurs when the receiver is over Alaska, as can be seen in Figure 13.8. Near the dropout, accuracies are not as good because the receiver is tracking the minimum four signals, and the dilution of precision is poor. These less accurate solutions can be detected and removed if necessary, at a reduced availability of total navigation solutions.

The estimated orbit can be compared to the precise orbit generated by the high-level GRACE system processing. These are the navigation solutions required for the extremely demanding gravity and altimetry primary mission objectives and can act as a “truth” reference for this exercise. A comparison between the least-squares estimated position generated using raw pseudoranges and the high-accuracy GRACE position output is shown in Figure 13.9. Figure 13.9 gives an intuitive sense of how a receiver’s position estimation accuracy can be improved with additional postprocessing.

It may seem curious that the dual-frequency solution is sometimes not as accurate as the single frequency solution. This is because each solution is computed with the maximum number of available measurements. The dual-frequency solution often uses less measurements because the semicodeless L2 measurements are weaker than the L1 measurements and sometimes are not available. Therefore the dual-frequency solutions experience worse DOP than the single-frequency solutions at

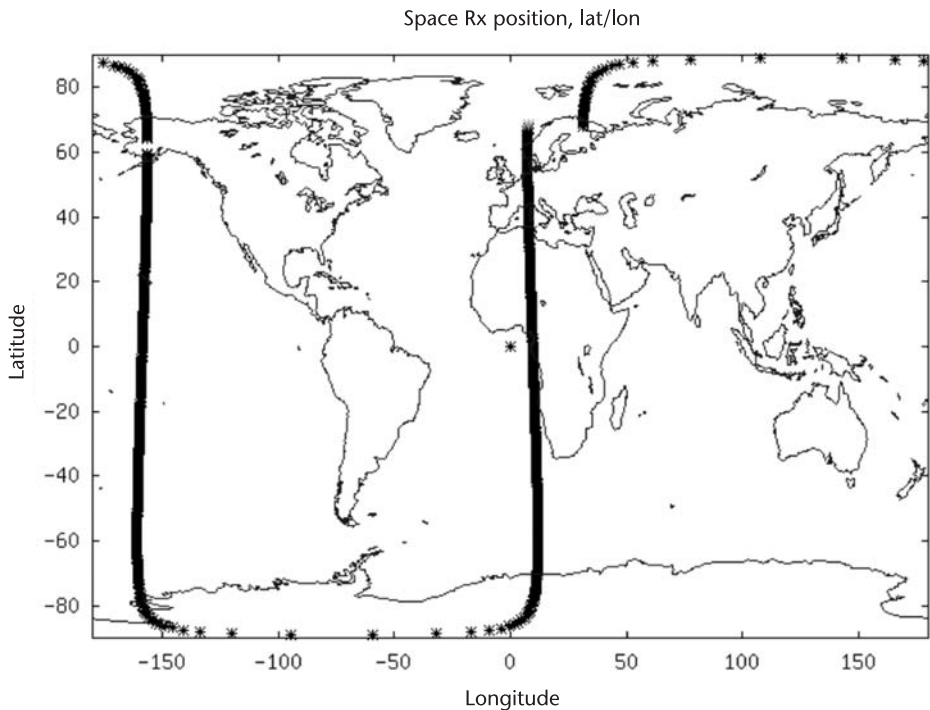


Figure 13.8 Latitude/longitude plot of the least-squares position estimates using raw pseudorange measurements over a single orbit on September 30, 2008.

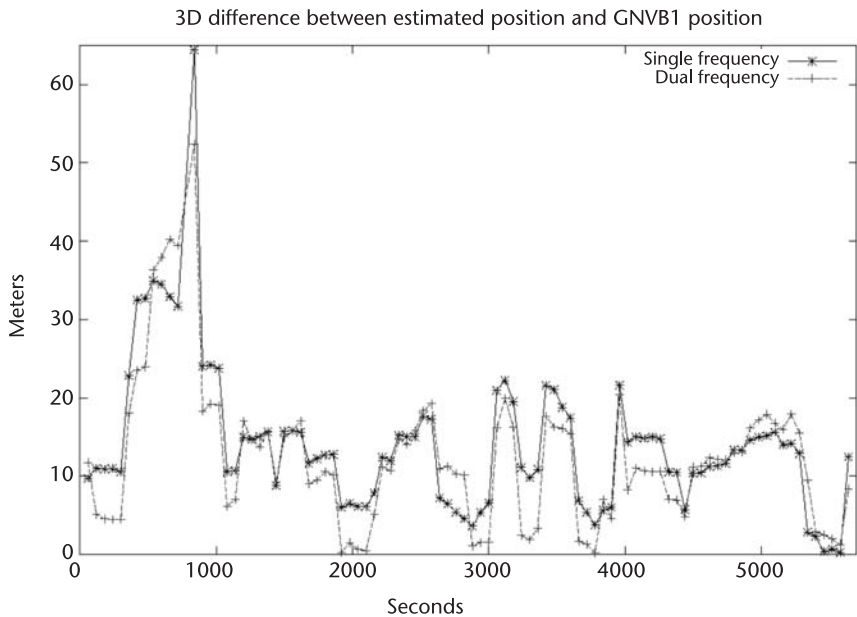


Figure 13.9 Difference between the least-squares position estimates calculated above and the high-accuracy GRACE position solution.

the same time instant. When the number of available measurements is the same, the dual-frequency solutions are more accurate.

13.6 Summary

With more than 25 years of practical experience in space, GNSS have already made important and diverse contributions to space systems and space science. The use of GNSS on numerous missions has led to a more focused understanding of the unique qualities of the space environment. Although there are difficulties that make the operating environment more challenging for space applications, the benefits of on-board autonomous navigation more than make up for these and have resulted in an overall advance in the capabilities of the space industry.

A large number of space applications have emerged and it is not possible to list them all in one chapter. However, certain navigation benchmarks may be used to characterize the state of the art. The most accurate orbit solutions available from GNSS in postprocessing have been verified at the centimeter level. By dynamically filtering dual-frequency GNSS measurements, half-meter real-time navigation solutions are achievable, depending on the application. Any single-frequency receiver that is capable of producing an on-orbit position fix should be able to attain standard positioning service accuracies of a few meters in real time.

It is now known that GNSS signals can be tracked at altitudes above the transmitting satellites with properly equipped receivers. GNSS measurements are also being used for special applications such as attitude determination and Earth science remote sensing. On-orbit autonomous operations such as satellite inspection and rendezvous should continue to develop over the next decade. GNSS metric tracking is being adopted as a standard for launch vehicles and ranges.

GNSS modernization provides a new set of measurements to space users and will likely lead to a renaissance in the development of new techniques and applications for GNSS services in space over the next quarter century.

References

- [1] Birmingham, W. P., B. L. Miller, and W. L. Stein, "Experimental Results of Using the GPS for Landsat-4 Onboard Navigation," *Navigation: Journal of The Institute of Navigation*, Vol. 30, No. 3, 1983, pp. 244–251.
- [2] Cohen, C. E., et al, "Space Flight Tests of Attitude Determination Using GPS," *International Journal of Satellite Communications*, Vol. 12, December 1994, pp. 427–433.
- [3] Winternitz, L., et al., "Navigator GPS Receiver for Fast Acquisition and Weak Signal Space Applications," *Proc. of the Institute of Navigation GNSS 2004 Conf.*, Long Beach, CA, Sept. 21–24, 2004, pp. 1013–1026.
- [4] Holt, G. N., and E. G. Lightsey, "In Situ Navigation of Spacecraft Formations in High Altitude and Extraterrestrial Orbits," *AIAA Journal of Spacecraft and Rockets*, Vol. 45, No. 2, March–April 2008, pp. 299–308.
- [5] Guzman, J. J., and A. Edery, "Mission Design for the MMS Tetrahedron Formation," *IEEE Aerospace Conference 2004 Proceedings*, Big Sky, MT, Mar. 6–13, 2004, pp. 533–540.

- [6] Borre, K., et al., *A Software Defined GPS and Galileo Receiver: A Single Frequency Approach*, Boston, MA: Birkhäuser, 2006.
- [7] Li, J., et al., "GPS Receiver Satellite/Antenna Selection Algorithm for the Stanford Gravity Probe B Relativity Mission," *Proc. of the 1999 National Technical Meeting of the ION*, San Diego, CA, Jan. 25–27, 1999, pp. 541–550.
- [8] Park, C. W., and J. P. How, "Quasi-Optimal Satellite Selection Algorithm for Real-Time Applications," *Proc. of the ION GPS-2001*, Salt Lake City, UT, Sept. 11–14, 2001, pp. 3018–3028.
- [9] Wu, S., et al., "MicroGPS for Orbit Determination of Earth Satellites," *Proc. of the 1996 National Technical Meeting of the ION*, Jan. 22–24, 1996, pp. 509–516.
- [10] Gaylor, D. E., E. G. Lightsey, and K. W. Key, "Effects of Multipath and Signal Blockage on GPS Navigation in the Vicinity of the International Space Station (ISS)," *Navigation: Journal of The Institute of Navigation*, Vol. 52, No. 2, 2005, pp. 61–70.
- [11] Bauer, F. H., et al., "The GPS Space Service Volume," *Proc. of the ION GNSS 2006 Conf.*, Fort Worth, TX, Sept. 26–29, 2006, pp. 2503–2514.
- [12] Wickert, J., et al., "GPS Radio Occultation with CHAMP and GRACE: Recent Results," in *Atmosphere and Climate Studies by Occultation Methods*, pp. 3–16, U. Foelsche, G. Kirchengast, and A. Steiner (eds.), Berlin: Springer-Verlag, 2006.
- [13] Moreau, M. C., et al., "Results from the GPS Flight Experiment on the High Earth Orbit AMSAT OSCAR-40 Spacecraft," *Proc. of the ION GPS 2002 Conf.*, Portland, OR, Sept. 24–27, 2002, pp. 122–133.
- [14] Powell, T. D., et al., "GPS Signals in a Geosynchronous Transfer Orbit: Falcon Gold Data Processing," *Proc. of the 1996 National Technical Meeting of the ION*, San Diego, CA, Jan. 25–27, 1999, pp. 575–585.
- [15] Ebinuma, T., et al., "A Miniaturized GPS Receiver for Space Applications," *16th IFAC Symposium on Automatic Control in Aerospace*, St. Petersburg, Russia, June 14–18, 2004.
- [16] Renaudie, C., et al., "Radiation Testing of Commercial-off-the-Shelf GPS Technology for use on Low Earth Orbit Satellites," *RADECS 2007, 9th European Conference Radiation and Its Effects on Components and Systems*, Deauville, France, Sept. 10–14, 2007.
- [17] Luthcke, S. B., et al., "The 1-Centimeter Orbit: Jason-1 Precision Orbit Determination Using GPS, SLR, DORIS, and Altimeter Data," *Marine Geodesy*, 26, pp. 399–421.
- [18] Yoon, Y., et al., "TerraSAR-X Precise Trajectory Estimation and Quality Assessment," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47, No. 1, 2009.
- [19] Montenbruck, O., and Ramos-Bosch, P., "Precision Real-Time Navigation of LEO Satellites Using Global Positioning System Measurements," *GPS Solutions*, 2008, 12, pp. 187–198.
- [20] Bertiger, W., et al., "GRACE: Millimeters and Microns in Orbit," *Proc. of the ION GPS 2002 Conf.*, Portland, OR, Sept. 24–27, 2002, pp. 2022–2029.
- [21] "Overview of the DART Mishap Investigation Results - For Public Release," NASA Marshall Space Flight Center, May 15, 2006.
- [22] Steyn, W. H., and Hashida, Y., "In-Orbit Attitude Performance of the 3-Axis Stabilized SNAP-1 Nanosatellite," *Proc. of the 15th Annual AIAA/USU Conference on Small Satellites*, Logan, UT, Aug. 13–16, 2001.
- [23] Gleason, S. T., *Remote Sensing of Ocean, Land and Ice Surfaces Using Bistatically Scattered GNSS Signals From Low Earth Orbit*, Ph.D. thesis, University of Surrey, United Kingdom, 2006.
- [24] Anthes, R. A., et al., "The COSMIC/Formosat-3 Mission: Early Results," *Bulletin of the American Meteorological Society*, Vol. 89, No. 3, 2008, pp. 313–333.
- [25] Gomez, S., "Three Years of Global Positioning System Experience on International Space Station," NASA/TP–2006–213168, August, 2006.
- [26] Freesland, D., et al., "GPS Based Attitude Determination-the REX II Flight Experience," *Proc. of the AIAA/USU Small Satellite Conference*, Logan, UT, September 16–19, 1996.

- [27] Wang, C., R. A. Walker, and M. P. Moody, "Single Antenna Attitude Algorithm for Nonuniform Antenna Gain Patterns," *AIAA Journal of Spacecraft and Rockets*, Vol. 44, No. 1, Jan.–Feb. 2007, pp. 221–229.
- [28] Bamford, W. A., et al., "A GPS Receiver for Lunar Missions," *Proc. of the 2008 National Technical Meeting of the ION*, San Diego, CA, Jan. 28–30, 2008, pp. 268–278.
- [29] Seok, B. S., et al., "Commercial GPS Receiver Design for High Dynamic Launching Vehicles," *Proc. of the 2004 International Symposium on GNSS/GPS*, Sydney, Australia, December 6–8, 2004.
- [30] National Research Council, *Streamlining Space Launch Range Safety*, National Academies Press, 2000.
- [31] Sega, R. M., "National Space Posture," Department of the Air Force Presentation to the House Armed Services Committee, United States House of Representatives, March 23, 2007.
- [32] Meehan, T. K., et al., "A Demonstration of L2C Tracking from Space for Atmospheric Occultations," *Proc. of the ION GPS 2008 Conf.*, Savannah, GA, Sept. 16–19, 2008, pp. 698–701.
- [33] Harris, R. B., Incorporation of the Global Positioning System Modernization Signals into Existing Smoother-based Ephemeris Generation Processes, Ph.D. thesis, The University of Texas at Austin, TX, 2008.
- [34] Lavrakas, J. W., "A Glimpse into the Future: A Look at GNSS in the Year 2017," *Proc. of the ION National Technical Meeting 2007*, Jan. 22–24, 2007, pp. 210–217.

Geodesy and Surveying

Chris Rizos and Dorota A. Grejner-Brzezinska

14.1 Introduction and Background

Since its introduction to the civilian community in the early 1980s the GPS has revolutionized almost all applications requiring high-accuracy positioning, navigation, or timing. Among the first civilian users, in fact, were the geodesists, who used GPS to: (1) observe the primary geodetic control networks that form the basis of all nations' map data, and (2) establish the global geodetic networks that define the international reference frame and monitor the motion of the crust with respect to that frame. Today GNSS continues to be the fundamental technology for these critical geodetic applications. However, as a result of progressive product improvements, GNSS technology is now also addressing the high-accuracy positioning requirements of many other nonscientific uses such as cadastral engineering, environmental planning, and geographic information system (GIS) surveys, as well as a range of new applications, such as precision location and guidance of machines, aircrafts, and ships. Furthermore, GNSS is being increasingly used for nonpositioning applications such as time transfer and atmospheric monitoring (see Chapter 15).

GPS/GNSS surveying is a differential (relative) technique (as described in Chapter 4) that can achieve centimeter-level accuracy using receivers that measure the carrier phase (CPH) of the underlying carrier wave signals. One receiver is located at a point of known coordinates—the *base station* or *reference receiver*—while the other receiver is operated by a user in such a way as to determine coordinates of one or more ground points, or the antenna's trajectory. Efficiencies in terms of time-to-first-fix (the length of time it takes the receiver to estimate its position after start up) and high-accuracy, are assured when all measurements are made on both the L1 or L2 carrier waves [1]. For baselines between GNSS receivers separated by more than, for example, 5–20 km (depending on the degree of ionospheric activity), it is necessary that dual-frequency (DF) receivers be used to correct for ionospheric delay on the signals (see Chapter 3). For shorter baselines, dual-frequency receivers are necessary for rapid initialization of centimeter-level accuracy positioning—ensuring robust, efficient precise positioning when enough GNSS signals are being tracked. Today CPH measurements to all visible GPS and GLONASS satellites are routinely used for GNSS geodesy and surveying—perhaps the first application of multiconstellation GNSS technology.

During the last decade and a half, GNSS surveying has benefited from a number of technological improvements that now enable centimeter-level accuracy positioning in real time, while the user receiver is in motion, over interreceiver distances of tens of kilometers, without the need for surveyors to operate their own base stations. These advances collectively have given rise to CPH-based positioning capability in the so-called real-time kinematic (RTK) mode that benefits from the establishment of permanent, continuously operating GNSS reference stations (see Section 14.3).

RTK and GNSS surveying applications are adequately addressed with reliable centimeter-level horizontal-positioning methods operating under local conditions. Geodetic applications, on the other hand, have been using GPS extensively since the early 1980s to address regional and global reference frame applications—these do not typically require real-time results, nor involve the user antenna being in motion [1]. GNSS (including GLONASS) geodesy has therefore mainly supported national and global datum definition and maintenance, as well as Earth science users, but at ever higher levels of accuracy. For example, *relative (differential) positioning* accuracies achieved today by geodetic techniques are typically at the *few parts per billion* (1 ppb is equivalent to millimeter accuracy for a thousand kilometer interreceiver separation).

14.1.1 GNSS Surveying

High-accuracy CPH-based positioning techniques have evolved since the early 1980s, from the first static geodetic control surveys to today's high efficiency GNSS-RTK techniques. The pseudorange-based DGPS or DGNSS techniques—developed in the 1990s primarily to address the policy of intentional standard positioning service accuracy degradation known as “*selective availability*”—are discussed in Chapter 4.

The characteristics of high-accuracy GNSS positioning techniques are summarized in Table 14.1 (adapted from Table 12.1, [2]). The following comments can be made:

- The least accurate differential positioning techniques are those based largely on pseudorange measurements (possibly with some CPH smoothing), in either single-base DGNSS mode or so-called “wide-area” DGNSS (WADGNSS) mode.
- The dual-frequency CPH-based techniques are superior in performance over single-frequency (SF) techniques, as they permit either (though not both at the same time) ionospheric measurement bias mitigation or very rapid CPH ambiguity resolution (see Section 14.2.1 for definition of CPH ambiguity).
- For all CPH-based positioning applications with relatively short static observation sessions (e.g., < 30 minutes), resolving the CPH ambiguities is crucial for centimeter-level accuracy. For static observation sessions of many hours in length, resolving ambiguities is not necessarily a prerequisite for centimeter-level accuracy—this is the typical geodetic mode of GNSS operations (Section 14.2.3).

Table 14.1 Current High-Accuracy GNSS Positioning Techniques (Adapted from Table 12.1 in [2])

<i>Range*</i>	<i>Accuracy[†]</i>	<i>CPH positioning type[‡]</i>	<i>Occupation time per point</i>	<i>Operational mode[§]</i>	<i>Infrastructure</i>
>100	0.5–2	WADGNSS [¶]	Instantaneous	RT, PR	SP
>100	0.05–0.5	Static (DF)	<30 minutes	PP, CPH+	CORS
>100	<0.05	Static (DF)-geodesy	Several hours	PP, CPH	CORS
<100	0.5–5	DGNSS	Instantaneous	RT, PR	SP
<100	<0.2	Static (DF)	>30 mins	PP, CPH	User
<50	0.05–0.5	Static (SF)	>30 mins	PP, CPH	User
<50	<0.01–0.2	Network-RTK	Near-instantaneous	RT, CPH	SP
<20	0.02–0.2	RTK	Near-instantaneous	RT, CPH+	SP or User
<20	<0.02	Fast-Static	<30 mins	PP, CPH+	CORS or User
<10	<0.02	RTK	Near-instantaneous	RT, CPH+	User

*Indicates distance from nearest reference station (limits are approximate only).

[†]95% confidence interval.

[‡]DF: Dual frequency; SF: single frequency.

[§]RT: Real-time with data link; PP: postprocessed; PR: pseudorange; CPH: carrier phase; CPH+: CPH plus PR.

^{||}SP: Service provider; CORS: continuously operating reference station; user: assumes users operate their own reference receiver.

[¶]WADGNSS: Wide Area Differential GNSS.

- High-productivity, high-accuracy (rapid GNSS positioning) requires comparatively short baselines (e.g., < 20km in length), and reliable CPH ambiguity resolution.
- All high-accuracy kinematic mode CPH-based GNSS positioning requires that the ambiguities be resolved, either before commencing the survey task or “on-the-fly” if ambiguities are “lost”—due to loss-of-lock or a cycle slip (see Section 14.2.2) on the tracked signals.
- The GNSS-RTK mode of operation is increasingly popular for *high-productivity* surveying, and time-critical applications such as machine guidance and navigation.
- The “single-base RTK” mode has been enhanced, from the late 1990s, by the so-called network-RTK approach, so that the spatially correlated atmospheric and satellite orbit errors can be better mitigated using *several* continuously operating GNSS reference stations that surround the area in which the user receiver is operating.
- The GPS datum is WGS84, while the GLONASS datum is PZ-90—or in English “Parameters of the Earth,” PE-90 [30]. For all but the most precise geodetic applications, WGS84 and PE-90 may be considered coincident with the *International Terrestrial Reference Frame 2005* [3].
- Reference receiver stations are operated under several modes: user, service provider, CORS. See Section 14.3 for a discussion on ground infrastructure to support high-accuracy GNSS surveying and geodesy.

The various modes of CPH-based positioning techniques used for *GNSS surveying* are described in Section 14.4.1.

14.1.2 GNSS Geodesy

According to Helmert’s classical definition, Geodesy is the “science of the measurement and mapping of the Earth’s surface” by direct measurements, such as terrestrial

triangulation, electronic distance measurement, leveling, and gravimetric observations. Additionally, in the past 50 years, this has expanded to include space techniques, based primarily on the tracking of a range of artificial Earth satellites. Following [4], the primary mission of Geodesy can be defined as:

- Establishment of the geodetic reference frame and the determination of precise global, regional and local 3D positions.
- Determination of the Earth's gravity field and its related models, such as the geoid, measurement and modeling of geodynamical phenomena, such as crustal deformation, polar motion, Earth rotation and tides.

The International Association of Geodesy (IAG – <http://www.iag-aig.org>) has established services for all the major *Satellite Geodesy* techniques: International GNSS Service (IGS – <http://www.igs.org>), International Laser Ranging Service (ILRS: <http://ilrs.gsfc.nasa.gov>), and the International DORIS Service (IDS: <http://ids.cls.fr>). These services generate “products” for a variety of users, including precise orbits, ground station coordinates, Earth rotation values, atmospheric parameters, and others. All have networks of ground tracking stations that form part of the physical realization of the International Terrestrial Reference Frame (ITRF) [3]. The definition of the ITRF requires the specification of the orientation and origin of its Cartesian axes, including the scale, physical constants and models used in its realization. For example, the size, shape and orientation of the reference ellipsoid that approximates the Earth's surface and the Earth's gravitational model are all precisely defined. The conventional realization of the ITRF is a set of coordinates and linear velocities (due mainly to crustal deformation and tectonic plate motion) of fundamental geodetic observatories (e.g., networks of stations of the IGS, ILRS, IDS), derived from satellite observations made at these continuously operating ground stations [5].

An increasingly important contribution of satellite geodesy is made by the GNSS satellites. Nowadays the majority of the fundamental stations underpinning the ITRF realization belong to the IGS network (Section 14.3.1). GNSS Geodesy therefore requires a considerable investment in fundamental tracking stations equipped with high-quality GNSS receivers, operating continuously over many years.

14.2 Technical Overview

This section introduces mathematical models for GPS observables and their processing strategies. One-way measurements and their corresponding observation equations are introduced first, followed by the differenced measurement models, a brief overview of the integer ambiguity resolution problem, and the data processing techniques for a single baseline and network-based approaches.

14.2.1 The Data Models and Processing Strategies of GNSS Geodesy and Surveying

Although the boundary is blurry, the difference between GNSS surveying and GNSS geodesy is primarily one of parameter accuracy. GPS geodesy nowadays requires

about 1,000 times greater relative positioning accuracy than GNSS surveying. Therefore, the major distinguishing characteristics of GNSS geodesy are the following (see also Section 14.2.3):

- It operates over long interreceiver distances, with long observation sessions—typically between 6–24 hrs.
- It is primarily concerned with computing relative positions within static networks of receivers for reference frame or crustal deformation monitoring, with subcentimeter levels of accuracy.
- It uses sophisticated data processing software that uses much more complex CPH mathematical models, and hence more estimable parameters, than GNSS surveying.
- It is increasingly addressing nonpositioning applications and supports other Earth science missions.

14.2.2 Mathematical Models

The majority of the high-accuracy GNSS techniques share a common requirement—application of the *differential* or *relative* positioning principle that requires one or more base stations or reference receivers operating at locations with known coordinates (see Chapter 4). This is in contrast to the standard single receiver *absolute* positioning mode of operation described in Chapter 3. (*Precise point positioning* is not yet a common high-accuracy CPH-based technique for GNSS surveying, though it will become increasingly more important for certain GNSS geodesy applications [6, 7].) The outcome of a GNSS survey task are coordinates relative to one reference receiver in the so-called single-base or baseline mode of positioning. In the case of user-operated reference receivers, the coordinates of the reference receiver define the datum in which the user receiver’s coordinates are expressed. GNSS geodesy, on the other hand, operates in the “multibase” or “network” mode. The resultant coordinates are in the network’s actual or implied datum. It is also possible to formulate mathematical models where the unknown (estimable) parameters are not coordinates (Section 14.2.4).

14.2.2.1 One-Way Observations

Pseudorange (PR) is the geometric range between the satellite transmitter and the receiver, distorted by the propagation media and the lack of synchronization between the satellite and the receiver clocks (see Chapter 3). It is recovered from the measured *time difference* between the epoch of the signal transmission and the epoch of its reception by the receiver. Currently, there are two types of pseudoranges: C/A-code PR and P-code PR (see Chapter 17 for a description of future signals that will produce additional measurements). Under the policy of “anti-spoofing” the P-code is encrypted to Y-code, resulting in more complicated signal recovery on the GPS L2 frequency. Since there is no C/A-code on the L2 carrier wave (except for the six Block IIR-M satellites launched in the past three years), signal correlation—as described in Chapters 2 and 5—is not normally possible, as most receivers cannot generate a replica of the unknown Y-code. Consequently,

more sophisticated—but less optimal—signal-tracking techniques must be used, and only geodetic/survey grade dual-frequency GPS receivers have L2 pseudorange measurement capability.

Dual-frequency pseudorange observations can be expressed as a function of the unknown receiver coordinates, satellite and receiver clock errors, and the signal propagation errors (maintaining notational consistency with equations in Chapter 3, and expanding the final error terms, ϵ^k):

$$\rho_{u,1}^k = r_u^k + c(\delta t_u - \delta t^k) + I_{u,1}^k + T_u^k + a_{u,2} + M_{u,1}^k + E_{u,1}^k \quad (14.1a)$$

$$\rho_{u,2}^k = r_u^k + c(\delta t_u - \delta t^k) + I_{u,2}^k + T_u^k + a_{u,3} + M_{u,2}^k + E_{u,2}^k \quad (14.1b)$$

where

r_u^k : The geometric distance between the receiver u and satellite k .

$\rho_{u,1}^k, \rho_{u,2}^k$: Psuedorange measured between receiver u and satellite k on L1 and L2.

I_1^k, I_2^k : Range error caused by ionospheric signal delay on L1 and L2.

δt_u : The u th receiver clock error.

δt^k : The k th satellite clock error.

c : The vacuum speed of light.

$a_{u,2}, a_{u,3}$: Interchannel bias terms for receiver u that represent the possible time nonsynchronization of the four measurements (CPH on L1 is considered as reference in the definition of interchannel bias used here).

$M_{u,1}^k, M_{u,2}^k$: Multipath on psuedorange observables at L1 and L2.

T_u^k : Range error caused by tropospheric delay between receiver u and satellite k .

$E_{u,1}^k, E_{u,2}^k$: Measurement noise for pseudorange on L1 and L2.

The position of a specific receiver $\mathbf{x}_u = [x_u, y_u, z_u]$ and the positions of the satellite at the time of transmission $\mathbf{x}^k = (x^k, y^k, z^k)$ express the true (or geometric) range as,

$$r_u^k = \sqrt{(x^k - x_u)^2 + (y^k - y_u)^2 + (z^k - z_u)^2} = \|\mathbf{x}^k - \mathbf{x}_u\| \quad (14.2)$$

PR measurements are used primarily for low-accuracy, *single-point*, or *absolute* positioning applications, as detailed in Chapter 3 where the satellite positions are often obtained from the broadcast ephemeris. However, they are also used *in combination* with CPH measurements to assist in ambiguity resolution. It is the CPH measurements made by the receiver to the tracked GNSS satellites that are used to obtain centimeter-accuracy positioning results.

Carrier phase is defined as a difference between the phase of the incoming carrier signal and the phase of the reference signal generated by the receiver. Since

at the initial epoch of the signal acquisition, the receiver can measure only the fractional phase, the CPH observable contains the initial unknown *integer CPH ambiguity* N . CPH ambiguity is the number of full phase cycles—of the L1 or L2 carrier wavelength—between the receiver and the satellite at the starting epoch, which remains constant as long as the signal tracking is continuous. After the initial epoch, the CPH-capable receiver can count the number of integer cycles that are being tracked. Thus, the CPH observable can be expressed as a sum of the fractional part φ (in cycles), measured with millimeter-level precision, and the integer number of cycles counted since the starting epoch t_0 . The integer ambiguity can be determined using special *ambiguity resolution algorithms*. Once the integer ambiguity is resolved, the ambiguous CPH observable can be converted to an *unambiguous* range measurement $R = (N + \varphi)\lambda$ by multiplying the sum of the measured phase (in cycles) and the initial integer ambiguity (in cycles) by the corresponding (L1 or L2) wavelength λ . The so-called phase-range observable Φ (in meters) equals the sum of R and all the error sources affecting the measurement (see Figure 14.1):

$$\begin{aligned}\Phi_{u,1}^k &= r_u^k + c(\delta t_u - \delta t^k) - I_{u,1}^k + T_u^k \\ &\quad + \lambda_1 N_{u,1}^k + \lambda_1(\varphi_{0,1}^k - \varphi_{u_0,1}) + m_{u,1}^k + e_{u,1}^k\end{aligned}\quad (14.3a)$$

$$\begin{aligned}\Phi_{u,2}^k &= r_u^k + c(\delta t_u - \delta t^k) - I_{u,2}^k + T_u^k \\ &\quad + \lambda_2 N_{u,2}^k + \lambda_2(\varphi_{0,2}^k - \varphi_{u_0,2}) + c_{u,2} + m_{u,2}^k + e_{u,2}^k\end{aligned}\quad (14.3b)$$

where:

$\Phi_{u,1}^k, \Phi_{u,2}^k$: Phase-ranges (in meters) measured between receiver u and satellite k on L1 and L2.

$\varphi_{0,i}^k, \varphi_{u_0,i}$: Initial fractional phases at the satellite (transmitter) and the receiver, respectively; i denotes the frequency.

$N_{u,1}^k, N_{u,2}^k$: Initial integer ambiguities on L1 and L2, corresponding to receiver u and satellite k .

$c_{u,2}$: Interchannel phase bias term between $\Phi_{u,1}^k$ and $\Phi_{u,2}^k$ $\lambda_1 \approx 19$ cm and $\lambda_2 \approx 24$ cm are wavelengths of the L1 and L2 carrier waves.

$m_{u,1}^k, m_{u,2}^k$: Multipath error on CPH observables on L1 and L2.

$e_{u,1}^k, e_{u,2}^k$: Measurement noise for CPH observables on L1 and L2.

Similar observations or mathematical equations can be developed for any of the GNSS signals. However, some of the frequency-dependent terms, as well as the noise and multipath terms, will be different. For example, in the case of GLO-NASS, the frequency of the carrier waves transmitted by each satellite is different [30].

Equation (14.2), which is the nonlinear part of (14.1) and (14.3), requires Taylor series expansion to enable the estimation of the three unknown user

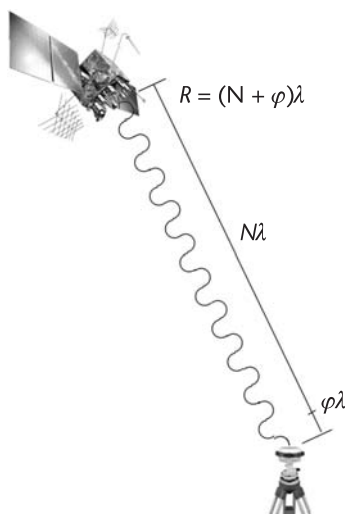


Figure 14.1 Carrier phase range measurement.

coordinates $(X, Y, Z)_r$. Secondary (“nuisance”) parameters in the above equations are satellite and receiver clock errors, tropospheric and ionospheric errors, multipath, broadcast ephemeris errors and integer ambiguities. These are usually removed or mitigated by (1) the differencing mode of GPS data processing (see next section); (2) empirical modeling (in the case of the troposphere); (3) processing of dual-frequency signals (in the case of the ionosphere); (4) estimation of extra parameters (in the case of the GNSS geodesy processing mode); (5) some combination of all of these.

14.2.2.2 Differenced Observations

A new PR or CPH observable is obtained by differencing the simultaneous measurements to the same satellites observed by the reference and the user receivers. Possible differencing methods include; “between-receiver differencing,” “between-satellite differencing,” or “between-epoch differencing.” The major advantage of relative positioning based on differenced measurements is the removal of the systematic error sources (common to the base station and the user, or both satellites and/or epochs of observation) from the resulting observable (see Table 14.2). This leads to increased positioning accuracy—from DGPS/DGNSS if using differenced PR measurements, or single-base or multibase techniques when using differenced CPH measurements.

For short to medium baseline lengths (up to ~20–50km) the systematic errors in GNSS observables due to the troposphere, satellite clock and broadcast ephemeris [contained within the geometric distance (14.2)] are of similar magnitude and spatially and temporally correlated. The differencing of CPH measurements allows for the removal—or at least a significant mitigation—of these systematic error sources.

The primary differencing operations are: (1) single-differencing mode, (2) double-differencing mode, and (3) triple-differencing mode. The differencing can be

Table 14.2 Differencing Modes and Their Error Characteristics

Error source	Single-Difference	Double-Difference
Ionosphere	Reduced, depending on the baseline length	Reduced, depending on the baseline length
Troposphere	Reduced, depending on the baseline length	Reduced, depending on the baseline length
Satellite clock	Eliminated	Eliminated
Receiver clock	Present	Eliminated
Broadcast ephemeris	Reduced, depending on the baseline length	Reduced, depending on the baseline length
Ambiguity term	Present	Present
Noise level with respect to one-way measurement	Increased by $\sqrt{2}$	Increased by 2

performed *between receivers*, *between satellites*, *between epochs* of observations, and even *between frequencies* (this observable does have some advantages for cycle slip detection, ambiguity resolution and ionospheric studies, but cannot be used for baseline processing). The single-differenced (between-receiver) measurement, $\Phi_{i,j}^k$, is obtained by differencing two observables to the satellite k , tracked simultaneously by two receivers i (reference) and j (user): $\Phi_{i,j}^k = \Phi_i^k - \Phi_j^k$ (see Figure 14.2).

By differencing observables from two receivers i and j , observing the same two satellites k and l , or simply by differencing two single differences to satellites k and l , one arrives at the double-differenced (between-receiver/between-satellite differencing) observable: $\Phi_{i,j}^{k,l} = \Phi_i^k - \Phi_j^k - \Phi_i^l + \Phi_j^l = \Phi_{i,j}^k - \Phi_{i,j}^l$ (see Figure 14.2). The *double-difference* (DD) is the most commonly used differential GNSS observable. Furthermore, differencing two double-differences, separated by the

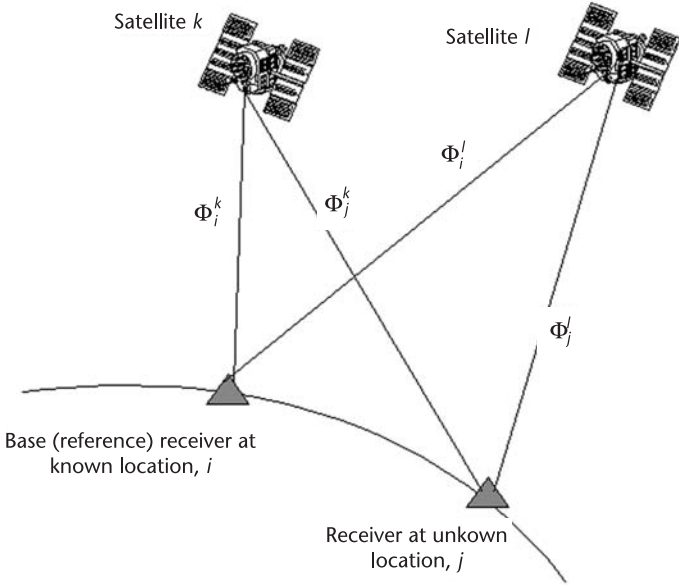


Figure 14.2 Between-receiver CPH double-differencing.

time interval $dt = t_2 - t_1$, generates the triple-differenced measurement, $\Phi_{i,j}^{k,l}(dt) = \Phi_{i,j}^{k,l}(t_2) - \Phi_{i,j}^{k,l}(t_1)$, which in the case of CPH observables effectively cancels the initial ambiguity term. Note that when four satellites are being tracked—the minimum number for three-dimensional GNSS positioning after integer ambiguities have been fixed—there are three independent DD observable combinations that can be formed from the eight one-way CPH measurements made by the two receivers to the four satellites. Table 14.2 shows the error characteristics for between-receiver single- and between-receiver/between-satellite double-differenced data [8]. Equations (14.4) and (14.5) present the final form of the double-difference PR and CPH equations, indicating that the error sources, as listed in Table 14.2, have been removed by the differencing operation.

$$\begin{aligned}\rho_{ij,1}^{kl} &= r_{ij}^{kl} + M_{ji,1}^{kl} + E_{ij,1}^{kl} \\ \rho_{ij,2}^{kl} &= r_{ij}^{kl} + M_{ji,2}^{kl} + E_{ij,2}^{kl}\end{aligned}\tag{14.4}$$

$$\begin{aligned}\Phi_{ij,1}^{kl} &= r_{ij}^{kl} + \lambda_1 N_{ij,1}^{kl} + m_{ji,1}^{kl} + e_{ij,1}^{kl} \\ \Phi_{ij,2}^{kl} &= r_{ij}^{kl} + \lambda_1 N_{ij,2}^{kl} + m_{ji,1}^{kl} + e_{ij,2}^{kl}\end{aligned}\tag{14.5}$$

14.2.3 Baseline Processing

In the case of GNSS surveying it is assumed that all systematic errors in the observations are either eliminated or are assumed to be so small in magnitude that the mathematical model of the basic DD-CPH observables only has two sets of estimable parameters: the Cartesian baseline components and the DD ambiguities. However, the degree of cancellation of the atmospheric (tropospheric and ionospheric) and orbital biases is dependent on the baseline length, since these errors tend to decorrelate with increasing receiver separation (Table 14.2). There are a number of comments that may be made with regards to the DD-CPH observable model:

- The *integer* nature of the GPS ambiguities is preserved after double-differencing—GLONASS processing is more complex and will not be dealt with here, though the reader is referred to [9, 10] for details.
- The DD ambiguities are assumed to be *constants* for as long as measurements are made in a continuous fashion over an observation *session*; hence *all* data contribute to the estimation of the unknown parameters. If, for whatever reason, the signal tracking to a satellite is broken, then the ambiguities involving that satellite before the break are *not equal* to the ambiguities after the break. That is, a *cycle slip* has occurred, and has to be “repaired” or otherwise accounted for as a new ambiguity term in the DD-CPH mathematical model.
- One set of coordinates in the observation model—corresponding to the reference receiver—is held fixed in the solution, while the second receiver’s coordinates are estimated. If the receivers are stationary (in the “static

- positioning” mode), the coordinates of the second receiver are unchanging from epoch to epoch, and hence all CPH data contributes to their estimation.
- This second receiver can be in motion (as in the case of “kinematic positioning”), requiring a new set of coordinates to be derived at each measurement epoch.
 - The entire computational process may be carried out in *postmission* mode, after the recorded measurements from both receivers are brought together within the baseline determination software. Alternatively, they may be computed in *real time*, if the reference station measurements are transmitted to the second receiver.
 - Commercial software is provided by the GNSS hardware manufacturers, and some third parties, for postprocessed and real-time modes of baseline estimation. Examples are: Trimble’s Geomatics Office, Topcon Tools, Leica’s Geo Office, and Javad’s Pinnacle.

14.2.3.1 Ambiguity Resolution

For centimeter-level positioning accuracy the initial DD ambiguity terms should be resolved to their integer values and then removed from the mathematical model for baseline processing. Carrier phase integer ambiguity resolution (AR) and validation is a complex problem that requires special algorithms and approaches (see paper by Donghuyn Kim and Richard Langley, gnss2000.kim.pdf, on the book DVD, or download from, <http://gauss.gge.unb.ca/papers.pdf/gnss2000.kim.pdf> for a review of AR techniques). There are three major categories of ambiguity resolution techniques, (1) AR in the measurements domain, (2) search technique in the coordinate domain, and (3) search technique in the ambiguity domain, which is the most accurate and widely used class of methods based on the integer least-squares theory (e.g., [11]).

The starting point for the ambiguity resolution (AR) process using category 3 methods is the real-valued estimates of the DD ambiguities derived from a least-squares solution using the simplified DD-CPH and DD-PR observation models. If these estimates are reliable then AR can be attempted with some degree of success. One of the most commonly used approaches to AR is the least-squares ambiguity decorrelation adjustment (LAMBDA) method. This technique is based on sequential conditional least-squares estimation preceded by a decorrelation of the ambiguities, by which the integer least-squares estimates can be computed very fast and efficiently [11].

In the early days of GPS positioning, one hour’s worth or more of CPH measurements were collected to ensure the separability of the baseline and ambiguity parameters within the least-squares solution. As the observation session lengthens, not only does the accuracy of the ambiguities increase but also the coordinate parameters. (In the limit, there would be no need to resolve the ambiguities because the coordinates would be determined to centimeter-level accuracy anyway—the reason why AR is not critical for GNSS geodesy.) Only with the drive to shorten observation sessions to a few tens of seconds or less (Section 16.4.1) did it become crucial that AR be carried out rapidly and reliably.

14.2.3.2 Baseline Determination Using Fixed Ambiguities

The basic steps in the baseline determination process are listed as follows [12].

1. Define the a priori values of the ambiguity parameters, generally from the least squares solution using the simplified DD-CPH and DD-PR model in (14.4-5). (For details on double differencing, the reader is encouraged to consult the good explanation in Pratap Misra and Per Enge's textbook [13]. Commonly, interfrequency linear combinations of L1 and L2 DD-CPH, such as, for example *widelane* ($\varphi_1 - \varphi_2$) combined with *narrow lane* ($\varphi_1 + \varphi_2$), L1 or L2 DD-CPH, are used, where φ_1 and φ_2 are the DD-CPH measurement on L1 and L2, respectively, measured in cycles.
2. Use a *search* algorithm to identify likely integer values for the ambiguities.
3. Employ a *decision-making* algorithm to select the “best” set of integer values.
4. Apply some *validation* tests to check whether these integer ambiguity values are indeed correct.
5. If the resultant ambiguity values are reliable, then convert the DD *ambiguous* CPH observables—using the “fixed” integer values—into DD *unambiguous* range observables—double-differencing $R = (N + \varphi)\lambda$ (14.5).
6. Compute the baseline parameters using the precise DD range observables.

In order to achieve centimeter-level coordinate accuracy in step (6), particularly for longer baselines, exceeding 5–10 km, the DD ionosphere-free carrier phase linear combination is formed to remove the differential ionospheric error, whose magnitude increases with the baseline length and the level of ionospheric activity. Note that (14.6) shows a generic form of the ionosphere-free linear combination, while in (14.7) it is assumed that the CPH measurements are in the DD mode.

$$\varphi_{1ion-free} = \alpha_1 \cdot \varphi_1 + \alpha_2 \cdot \varphi_2 \quad (14.6)$$

$$\varphi_{1ion-free} = \frac{f_1}{c} \cdot r + \alpha_1 \cdot N_1 + \alpha_2 \cdot N_2 \quad (14.7)$$

where:

φ_1 is the CPH measurement on L1 [in cycles] with ambiguity N_1 ;
 φ_2 is the CPH measurement on L2 [in cycles] with ambiguity N_2 ;
 $\varphi_{1ion-free}$ is the ionosphere-free CPH linear combination [in cycles], here

$$\alpha_1 = \frac{f_1^2}{f_1^2 - f_2^2} \approx 2.546$$

$$\alpha_2 = \frac{-f_1 f_2}{f_1^2 - f_2^2} \approx -1.984$$

f_1 is the L1 frequency;

f_2 is the L2 frequency;

c is the vacuum speed of light.

Alternatively, if (14.6) and (14.7) are scaled by the L2 frequency, the ionosphere-free L2 carrier phase measurement is obtained, having the same form as (14.6) except that α_1 would be replaced by β_1 , and α_2 replaced by β_2 , where

$$\beta_1 = \frac{f_1 f_2}{f_1^2 - f_2^2} \approx 1.984 \quad \text{and} \quad \beta_2 = \frac{f_2^2}{f_1^2 - f_2^2} \approx -1.54. \quad \text{The ambiguity term of the}$$

DD ionosphere-free linear combination is no longer an integer, being $N_{1ion-free} \approx 2.546N_1 - 1.984N_2$ (when expressed in units of L1 cycles) or $N_{2ion-free} \approx 1.984N_1 - 1.54N_2$ (when expressed in units of L2 cycles).

AR *reliability* [steps (3)–(5) above], in the context of GNSS surveying, is a function of the following:

- Baseline length (the shorter the better);
- The number of satellites (the more the better);
- Whether satellites rise or set during the session (continuous tracking of satellites across the entire observation session is preferred);
- The satellite-receiver geometry [low DOP values (see Chapter 3)];
- The degree of multipath disturbance to the measurements (the less the better);
- Whether observations are made on both carrier frequencies (it is much easier to resolve ambiguities when dual-frequency observations are used);
- The length of the observation session (the longer the better).

Although the above remarks imply that AR is a process that can only be applied to static GPS CPH data, during the last decade and half the AR algorithms have been significantly refined and the receiver hardware improved to such a point that AR can be carried out using just a few tens of seconds of tracking, even if the receiver is in motion—known as “on-the-fly” AR. GNSS is now a very powerful surveying technology; however future multiconstellation GNSS will perform even better than today’s GPS+GLONASS systems (Section 14.5.2).

14.2.4 Network Processing for Positioning

In contrast to the baseline mode of positioning using a simplified DD-CPH mathematical model, in the case of GNSS geodesy it cannot be assumed that systematic error sources are eliminated or mitigated by double differencing. Hence the mathematical model of DD-CPH contains additional estimable parameters for *residual* satellite orbit error, tropospheric delay, and several other nuisance effects that cannot be ignored for very high-accuracy GNSS. The following are some characteristics of data processing and operations for very high-accuracy positioning applications:

- The interreceiver distances are far greater than is typical in GNSS surveying, ranging from tens of kilometers up to global scales of thousands of kilometers.
- It may not be necessary to resolve (and fix) the integer double-differenced ambiguities if long observation sessions are used (up to 24 hours in length).

- All data is typically archived in receiver independent exchange (RINEX) format (see DVD for copy of manual or download from, <ftp://epncb.oma.be/pub/data/format/rinex300.pdf>), and postprocessed.
- All data is processed using scientific software packages that have been developed over many years by researchers—including GAMIT/GLOBK (from MIT/Scripps Institution of Oceanography), Bernese (from the University of Bern), GIPSY (from JPL), PAGES (from National Geodetic Survey). These are examples of stable, well-maintained and freely available processing software packages.
- The static mode of observation session processing is typically used, though ground station velocities may also be estimated from multiday processing.
- Data from many simultaneously tracking receiver stations may be processed together—perhaps from over 100 GNSS receivers—in network processing mode.
- Multiday and multinetwork processing can be carried out in a subsequent step where intermediate solution output in the form of SINEX files (see DVD for copy of manual or download from, <http://www.iers.org/MainDisp.-csl?pid=190-1100110>) are combined, and the geodetic parameters of interest estimated.
- A priori satellite orbits (and perhaps other parameters such as tropospheric and ionospheric delays) typically are obtained from the *International GNSS Service* (Section 14.3.1).
- The geodetic applications are myriad, and include nonpositioning applications.

With the expansion of networks of fixed, permanently operated reference receivers (Section 14.3), the capabilities of GNSS geodesy are increasing—making it an indispensable space geodesy technology.

14.3 GNSS Ground Infrastructure—Continuously Operating Reference Station (CORS) Networks

While the user-operated reference station is still a common mode of GNSS surveying for baseline determination, increasingly many classes of users are taking advantage of *ground infrastructure* of permanent continuously operating reference stations (CORS).

CORS nowadays provide the fundamental infrastructure required to meet the needs not only of geodesy and the geosciences, but also of many professional GNSS surveying, mapping, and navigation users. Furthermore, the widespread use of the GNSS-RTK technique means that such reference station receivers increasingly have to support ever expanding nongeodetic, *real-time* applications of high-accuracy positioning for engineering, machine guidance, and precision agriculture. This requires additional investment in communication links—between reference receivers for monitoring performance and computing network parameters, as well as with real-time users via wireless telecommunications. The following discussion has been adapted from [14].

GPS in the 1980s was almost exclusively used for geodetic control surveys, and the interreceiver distances were at first several tens of kilometers, being the average distance between first order geodetic groundmarks. However, at about this time GPS was also proving itself to be an effective tool of *space geodesy* for measuring crustal motion and establishing the global reference frame. Subsequently, the distances between receivers increased progressively to hundreds and then thousands of kilometers, while the relative accuracies simultaneously increased. These developments ensured centimeter-level accuracy within GPS receiver networks even as interreceiver distances grew significantly. These GPS geodetic stations inevitably became permanent reference stations for: (1) the monitoring of the station motion itself (due to horizontal and vertical crustal motion), (2) defining modern geocentric geodetic datums at the national level, and (3) the extension and increasing density of the geodetic control (groundmark) networks using GPS techniques.

14.3.1 The IGS Infrastructure

It is important to recognize the significant contribution of the “super-network” of reference stations of the *International GNSS Service* (IGS) to Geodesy, and to the GNSS community in general. Several hundred globally distributed GPS receivers (many now with GLONASS tracking capability) operate on a continuous basis, many for over 10 years, contributing data to the various IGS analysis centers and other users [15, 16]. Notably, tools for processing IGS ephemeris files have been included on the DVD as part of Chapters 5 and 13.

14.3.1.1 The IGS Today

The IGS was established in January 1994 as a service of the International Association of Geodesy (IAG) [16]. Since June 1992 the IGS—originally known as the “International GPS Service for Geodynamics,” from 1999 simply as the “International GPS Service,” and finally since March 2005 as the “International GNSS Service”—has been making freely available to all users raw GNSS tracking data from its global network, and high-accuracy satellite ephemerides and other derived products (Table 14.3).

The mission of the IGS is “to provide the highest-quality GNSS data and products in support of the terrestrial reference frame, Earth rotation, Earth observation and research, positioning, navigation and timing and other applications that benefit society” [16]. The IGS activities are fundamental to scientific disciplines concerned with climate, surface weather, sea level change, gravity, space weather research, and more. However, the IGS also supports many other applications including precise navigation, machine automation, time transfer, and surveying and mapping.

The IGS operates as a voluntary, noncommercial, confederation of about 200 institutions worldwide (Figure 14.3), self-governed by its members, managed on a day-to-day basis by the Central Bureau, under the policy guidance of its Governing Board.

The IGS collects, archives, and distributes GPS and GLONASS observation data sets of sufficient accuracy to meet the objectives of a wide range of scientific

Table 14.3 The IGS Product Summary [15, 16]

		Accuracy	Latency	Updates	Sample Interval
<i>GPS Satellite Ephemerides/Satellite and Station Clocks</i>					
Broadcast	Orbits	~160 cm	Real time	—	Daily
	Sat. clks	~7 ns			
Ultrarapid	Orbits	~10 cm	Real time	Four x daily	15 min
(predicted half)	Sat. clks	~5 ns			
Ultrarapid	Orbits	<5 cm	3 hours	Four x daily	15 min
(observed half)	Sat. clks	~0.2 ns			
Rapid	Orbits	<5 cm	17 hours	Daily	15 min
	Sat. & Stn. clks	0.1 ns			5 min
Final	Orbits	<5 cm	~13 days	Weekly	15 min
	Sat. & Stn. clks	<0.1 ns			5 min
<i>GLONASS Satellite Ephemerides</i>					
Final		15 cm	2 weeks	Weekly	15 min
<i>Geocentric Coordinates of IGS Tracking Stations (>130 sites)</i>					
Final positions	Horizontal	3 mm	12 days	Weekly	Weekly
	Vertical	6 mm			
Final velocities	Horizontal	2 mm/yr	12 days	Weekly	Weekly
	Vertical	3 mm/yr			
<i>Earth Rotation Parameters</i>					
Ultrarapid	PM ¹	0.3 mas	Real Time	Four x daily	Four x daily
(predicted half)	PM rate	0.5 mas/day			(00, 06, 12, 18 UTC ³)
	LOD ²	0.06 ms			
Ultrarapid	PM	0.1 mas	3 hours	Four x daily	Four x daily
(observed half)	PM rate	0.3 mas/day			(00, 06, 12, 18 UTC)
	LOD	0.03 ms			
Rapid	PM	<0.1 mas	17 hours	Daily	Daily (12 UTC)
	PM rate	<0.2 mas/day			
	LOD	0.03 ms			
Final	PM	0.05 mas	~13 days	Weekly	Daily (12 UTC)
	PM rate	<0.2 mas/day			
	LOD	0.02ms			
<i>Atmospheric Parameters</i>					
Final tropospheric zenith path delay		4 mm	<4 weeks	Weekly	2 hours
Ultra-Rapid tropospheric zenith path delay		6 mm	2–3 hours	Every 3 hours	1 hour
Final ionospheric TEC ⁴ grid		2-8TECU ⁵	~11 days	Weekly	2 hours; 5 deg (lon) x 2.5 deg (lat)
Rapid ionospheric TEC grid		2-9TECU	<24 hours	Daily	2 hours; 5 deg (lon) x 2.5 deg (lat)

Notes: ¹PM – Polar Motion; ²LOD – Length of Day; ³UTC – Universal Time Coordinated; ⁴TEC – Total Electron Content; ⁵ TECU – TEC units

and engineering users. These data sets are analyzed and combined to form the IGS products shown in Table 14.3 [16]. While the IGS product range has been mainly concerned with GPS, since 1998 GLONASS products were also developed, initially in connection with the International GLONASS Experiment of 1999. This continued seamlessly, from 2001, through the International GLONASS Service, a pilot project of the IGS, which reached a successful conclusion in December 2005—and thus could be dissolved—when the GLONASS products (raw data and derived products) were integrated into the mainstream IGS product flow (see Table 14.3).

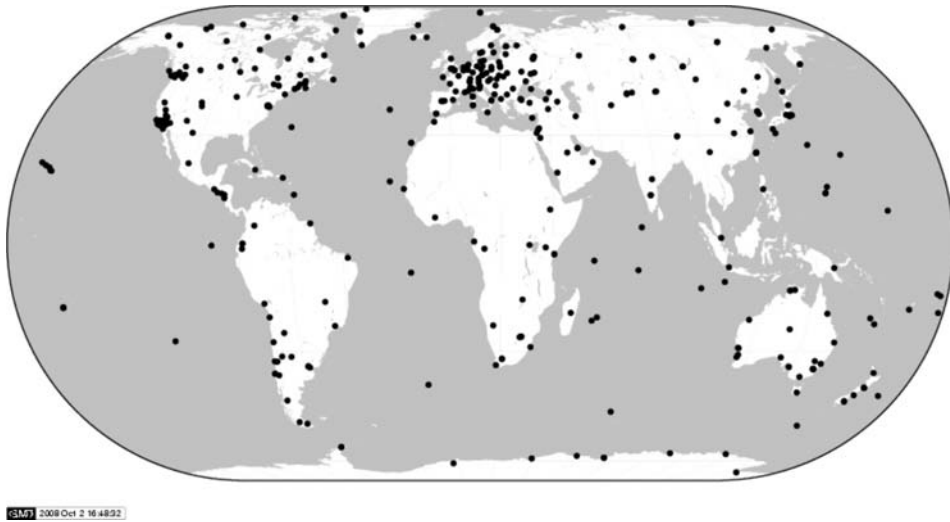


Figure 14.3 Global distribution of IGS stations. Image taken from <http://igsceb.jpl.nasa.gov/network/netindex.html>].

14.3.1.2 The IGS in the Future

The expression “GNSS system of systems” is now increasingly used for receiver hardware, or data processing software, or CORS infrastructure (see Sections 14.3.2 and 14.5.2) that can track signals, process measurements or provide user services from the different GNSS constellations, which will be operational during the next decade—GPS, GLONASS, GALILEO, COMPASS (see Chapter 17). It is worth speculating on what a future “GNSS system of systems” CORS infrastructure may consist of. Options could include [17],

1. True “GNSS system of systems.” CORS receivers able to track *all* GNSS signals—across four distinct L-band frequencies—will likely be established with relatively large interreceiver separations, perhaps of the order of several hundred kilometers or more. These may comprise the future backbone of the IGS.
2. Lower cost multiconstellation GNSS CORS receivers, probably with dual-frequency tracking capability, will likely be established at closer receiver spacings, from just several kilometers apart (to support structural deformation monitoring, and single-base RTK applications) and up to several tens of kilometers (to support most DGNSS and network-RTK users). The interoperable frequencies are most probably the L1 (1575.42-MHz) and L5 (1176.45-MHz) frequencies, which all GNSS are likely to be transmitting.

It should be noted that all, or even most, of the signals from a “GNSS system of systems” will not be transmitted before 2013 at the earliest, hence the current investment in CORS infrastructure for the IGS (and national CORS networks) will continue to be in GPS+GLONASS capable receivers. However, the upgrade of the CORS infrastructure at or after 2014 will have to incorporate multiconstellation GNSS tracking capability.

In 2008 the IGS launched a Real-Time Pilot Project [18], with the following objectives [16]:

- Implement a global IGS real-time (RT) GNSS tracking network;
- Enhance and improve selected IGS products;
- Generate new RT products;
- Investigate standards and formats for RT data collection, data dissemination, and delivery of derived products;
- Monitor the integrity of IGS predicted orbits and GNSS status;
- Distribute observations and derived products to RT users;
- Encourage cooperation among RT activities, particularly in IGS densification areas.

If successful, new RT IGS products will be generated on a continuous basis, to augment the current set of postprocessed products (Table 14.3). In parallel with this internal development, the IGS is working with its parent organization—the IAG—on the design of the *Global Geodetic Observing System* (GGOS) (<http://www.ggos.org/>). GGOS will integrate the activities and products of the IAG services and commissions, to provide the geodetic component of the Global Earth Observing System of Systems (GEOSS) now being established by the intergovernmental Group on Earth Observations (GEO) (<http://www.earthobservations.org/>). This contribution will be vital to *Global Change* studies, and necessitates an upgrade of the global geodetic capability. The IGS, with its prime concern for high-accuracy and high-reliability processing of the signals of the GNSS constellations, and as provider of the consolidated inputs of the GNSS contribution to the ITRF, will play a key role in GGOS.

To satisfy these internal (RT services) and external (GGOS) objectives, elements of the IGS global tracking network (currently numbering several hundred CORS) will need to be upgraded. New GNSS receivers, stable monumentation, and improved communications links will be required—as will improved modeling and data processing methodologies. Such a CORS upgrade strategy will no doubt also be followed by different countries with respect to their own CORS networks.

14.3.2 National CORS Infrastructure

While the IGS network is an example of a globally distributed Continuously Operating Reference Station network, many CORS networks have been established at national, state, and even local levels. This trend of establishment of CORS will continue into the future, and will evolve to handle different GNSS signals as they start to be transmitted over the next 5–10 years.

14.3.2.1 CORS Today

In the 1990s, when the establishment of CORS networks was justified on geodetic grounds, national networks were similar to IGS stations. That is, although operated on a “24/7” basis, the networks only periodically downloaded data from each

receiver as RINEX files and sent it on to an archive or data center. From there the GNSS observation data was available to users for postprocessing.

Archived RINEX files from both IGS stations and national GNSS CORS networks are now accessed by users via the Internet. All IGS data has been, and will continue to be, available free-of-charge to all users. Despite the fact that some GNSS receiver network operators charge fees for their RINEX files, the increasing trend is to make such data available at no cost. If users were: (1) satisfied with postprocessed results (i.e., they do not want coordinates in real time), and (2) were fortuitously carrying out a GNSS positioning task *close* to a CORS, then users could benefit from such data in either of two ways:

- They could download data from the nearest GNSS CORS, for the time period of their own survey, and then process this data together with their receiver's measurements using their own software.
- Alternatively there are several free “Web engines” that accept data uploaded by a user, combine it with nearby IGS station data, and carry out the data processing for them (Section 14.4.2).

The definition of “close” in the case of centimeter-level accuracy applications depends on the GNSS technique and operational mode that is used [1], ranging from no more than approximately 10 km for “rapid-static” or “on-the-fly” CPH-based techniques using off-the-shelf commercial software, to perhaps a few hundred kilometers if scientific software is used.

With the advent of GPS-RTK techniques in the early 1990s (and more recently GNSS-RTK, incorporating GLONASS CPH data), CPH-based positioning finally could be seriously considered a *surveying tool*. *Productivity* has increased to such a degree that private survey companies now invest significantly in the receiver equipment. (*Productivity* can be measured in many ways, but essentially refers to the number of ground points that can be coordinated in a work day, with minimum operational constraints.)

However, to ensure high productivity GNSS-RTK there are several constraints, including: (1) that all GNSS receivers must have dual-frequency tracking capability and (2) that the interreceiver distances should be less than 10 or so kilometers. These operational constraints placed cumbersome limits on many applications and contributed to the development of *network-based techniques* [7, 19–23] that have enabled centimeter-accuracy positioning with less dense reference receiver spacing requirements (on the order of 50–100 km) and that could operate in real time. Such relaxed specifications on CORS receiver spacing has encouraged many more countries to establish real-time CORS networks—as an investment in *surveying infrastructure* (Figure 14.4).

14.3.2.2 Next-Generation CORS

Next-generation GNSS, with many more signals and frequencies, will have important implications for the development of CORS networks. For example, with the benefits of extra satellite visibility (or availability), and CPH tracking of three or more different transmitted L-band frequencies, the interreceiver separations

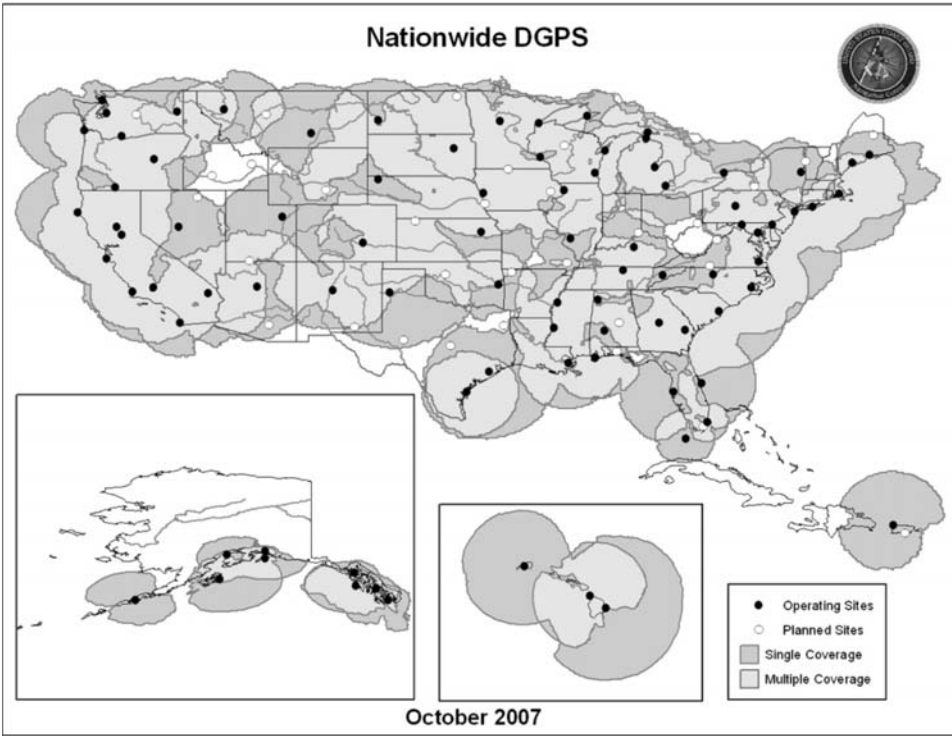


Figure 14.4 U.S. DGPS coverage (<http://www.navcen.uscg.gov/dgps/default.htm>).

within GNSS CORS networks can be relaxed even further. Permanent CORS networks are increasingly facilitating real-time techniques such as GNSS-RTK, and this trend will become more evident with future CORS capable of multi-GNSS tracking. For instance, “single-base RTK” will be possible over baseline lengths that are over a hundred kilometers with centimeter-level accuracy, albeit with lowered reliability vis-à-vis “network-RTK” techniques. Furthermore, if decimeter-level coordinate accuracy is adequate, the CORS separation can be relaxed to several hundreds of kilometers, significantly reducing the necessary ground infrastructure investment required.

No discussion on future CORS infrastructure would be complete without addressing the question: “Who will establish and maintain future CORS networks, government or private service providers?” Government agencies and organizations typically justify the costs of implementing CORS networks by citing the principle of “preventable costs.” This is similar to the strategy used to finance the establishment of classical geodetic groundmark networks decades earlier. The return on the original investment is not measured in terms of revenue earned, but justified as a means of keeping the costs borne by the community lower than the alternative (i.e., having no geodetic control infrastructure). This approach also encourages network standardization and avoids the establishment of a patchwork of private, ad-hoc networks for project-specific purposes. The work undertaken in [14] identified the following models for CORS networks, and services derived from them:

1. Institutional CORS infrastructure with no commercial services;
2. Government CORS infrastructure, which provides the foundation for commercial services;
3. Government CORS infrastructure combined with licensing the data to the private sector;
4. Cooperative privately owned CORS infrastructure that operates commercial services;
5. Privately owned CORS infrastructure that operates commercial services.

Which of these models—or combination/variations of the above—will prevail in different countries in an era of “GNSS system of systems,” and an expanding high-accuracy user base, is impossible to answer at this point in time. There are already several different *business models* for CORS services (real-time positioning, or raw GNSS measurements) in the United Kingdom, Germany, the United States, Japan, and elsewhere. However, in many countries the trend is increasingly towards option (3), where a private company *wholesales* GNSS CORS data to service providers—who then market derived GNSS-RTK services to certain market segments—or *retails* the data/service directly to customers. In addition, the technical specifications of future “GNSS system of systems” CORS receivers are still unresolved (Section 14.5.2).

14.4 Surveying and Geodesy Applications and Operational Modes

There are a number of criteria that can be used to categorize GNSS techniques, operational modes and applications. Consider, for example, the following:

1. *Static GNSS positioning techniques*: High-accuracy relative positioning techniques based on CPH measurements from *stationary* GNSS user and reference receivers, typically in postprocessing mode;
2. *GNSS nonpositioning uses*: For such functions as timing, atmospheric remote sensing, and other geodetic applications;
3. *Kinematic GNSS positioning techniques*: Versatile relative positioning techniques that use CPH measurements to determine the coordinates of a *moving* user GNSS receiver, including in real time.

The first category can be further broken down into surveying techniques on the one hand, and those that are used for geodesy applications. The distinguishing characteristics of the two classes of static positioning techniques are listed as follows:

- Medium-to-short baseline GNSS survey techniques requiring relative accuracies at the few parts per million level for baselines typically <50 km in length, using short observation sessions, with data processing (in real time or postprocessed mode) using commercial software packages;

- Ultra-precise, long baseline GNSS techniques involving accuracies from a few parts per million to several parts per billion, characterized by top-of-the-line geodetic GNSS receivers and antennas, many hours (and even days) of observations, and data processing with sophisticated scientific software.

The second category of GNSS nonpositioning applications are briefly discussed in Section 14.4.2. The third category, the high-accuracy kinematic GNSS positioning techniques, are considered in Section 14.4.1 within the class of GNSS surveying—although the technique is also sometimes referred to as “high-precision navigation.”

14.4.1 GNSS Surveying

The main weakness of conventional GPS/GNSS surveying is that the observation session lengths are comparatively long, the results are obtained well after the survey (i.e., postmission), and the field procedures are rigid. During the late 1980s considerable attention was paid to these issues as they were considered to be unnecessarily restrictive. That is, if antennas could be moving during a survey, then new applications using GNSS technology could be addressed. If the length of time required to collect CPH data for centimeter-level accuracy solutions could be reduced, then survey productivity would improve and the technology would be attractive for more surveying applications. If the results could be obtained immediately after the measurements were made, then the technology could be used for time-critical missions such as engineering stakeout and construction machine guidance. *GNSS-RTK is therefore one of the most important innovations in CPH-based GNSS positioning because it has opened up high-accuracy positioning to many more user groups.*

14.4.1.1 Applications

Generally, GNSS surveying must satisfy high-accuracy positioning requirements under a variety of operational conditions, such as: (1) on land, at sea, and from the air; (2) for both stationary (static) and moving (kinematic) applications; (3) with relatively high “productivity”; (4) a minimum of field constraints; and (5) increasingly, in real time. Some example applications include the following:

1. *Land surveys:* Establishment of minor geodetic control for other survey tasks; engineering construction and “as-built” surveys; topographic or “detail” surveys including GIS data capture, digital terrain modeling and architectural surveys; georeferencing points-of-interest for a variety of environmental mapping applications; road centerline surveys; deformation surveys (dams, buildings, bridges, and other structures); cadastral or legal boundary surveys for land subdivision, right-of-way acquisition, etc.; open-cut mining operations; geophysical surveys; machine automation applications in engineering, mining, agriculture, and robotics.
2. *Marine surveys:* Surveys of hydrographic vessels in harbors and waterways, as well as in the open ocean, in support of chart preparation (georeferencing

depth sounding and side-scan sonar data), pipeline or cable-laying operations and dredging; precise harbor navigation including automated ship berthing, traversing shallow channels and real-time ship-keel clearance determination; ocean tide buoy tracking; geophysical surveys.

3. *Airborne surveys*: Determination of aircraft trajectories in support of photogrammetric operations (“aerial mapping”), airborne laser scanning, or LADAR as described in Chapter 8 and geophysical surveys such as aeromagnetism, gravimetry, hyperspectral; and increasingly for specialized UAV operations, “formation-flying.”

14.4.1.2 GNSS Surveying Techniques: General Remarks

Whether the application is operating in GNSS-RTK or postprocessing mode, a further generalization can be made: The deployment of *high-productivity* CPH-based techniques is one of the most important distinguishing characteristics of GNSS surveying. These modern GNSS surveying techniques are given a variety of names by the different GNSS receiver manufacturers, but the following generic terms can be used [24]:

- *Rapid-static* positioning technique;
- *Stop-and-go* technique;
- *Kinematic* positioning technique.

All require the use of specialized hardware and software, as well as specialized field procedures. It is important to emphasize that GNSS receivers capable of executing these types of surveys can also be used for conventional static GNSS geodesy. That is, it is *not* the GNSS instrumentation that distinguishes surveying from geodesy. Each of the above techniques is a technological solution to the challenge of ensuring high productivity and operational flexibility without sacrificing very much in terms of *accuracy* and *reliability*. However, none of these techniques is as accurate or reliable as conventional static CPH-based GNSS positioning, and because each of these techniques has unique advantages they are all typically offered as positioning options within modern GNSS products.

14.4.1.3 Rapid-Static Technique

The field procedures are much like those for conventional static GNSS techniques except that: (1) the station occupation times are shorter, (2) the baselines should be comparatively short, (3) the satellite geometry favorable, and (4) signal disturbances such as multipath should be kept to a minimum [25]. It is currently not possible to define exactly how much data needs to be collected in order to produce quality baseline solutions every time, based on resolved CPH ambiguities. However, equipment user manuals typically give guidelines in this regard. Some receivers also provide an audio and/or visual *indication* when enough data has been collected in the field (but this cannot be confirmed until the measurement data is downloaded and processing is completed). If the RTK positioning mode is employed then this

data quantity “gamble” can be overcome as the user equipment indicates when it has generated ambiguity-fixed solutions in the field.

The National Geodetic Survey operates the On-line Positioning User Service (OPUS) as a means to provide GPS users easier access to the National Spatial Reference System (NSRS). OPUS-RS [26] is a new version of OPUS designed to obtain geodetic quality positioning results from user data sets as short as 15 minutes, using national CORS (see, http://www.ngs.noaa.gov/OPUS/What_is_OPUS-RS.html). OPUS-RS searches the CORS network in order of increasing distance from the user’s station, selecting six reference stations that have suitable data and are within 250 km from the user. Because of the geographic interpolation algorithms used, OPUS-RS will only process a file for which the user’s station is either inside the polygon enclosing the selected reference stations, or no more than 50 km outside that polygon. This restriction can be overridden if the user manually selects reference station(s) using the OPUS Options page. The results are e-mailed to the user. Centimeter-level accuracies from as little as 10 minutes of data (see Figure 14.5) were reported by [27, 28]. Example RINEX data files suitable for processing by OPUS have been included on the DVD.

Example AR statistics are listed in Table 14.4 for a rapid-static solution supported by three CORS stations separated from the user by 135–162 km [27].

The rapid-static technique is well-suited for short-range applications such as control densification and engineering surveys, or any job where many ground points need to be coordinated. Unlike the *kinematic* and *stop-and-go* techniques, there is

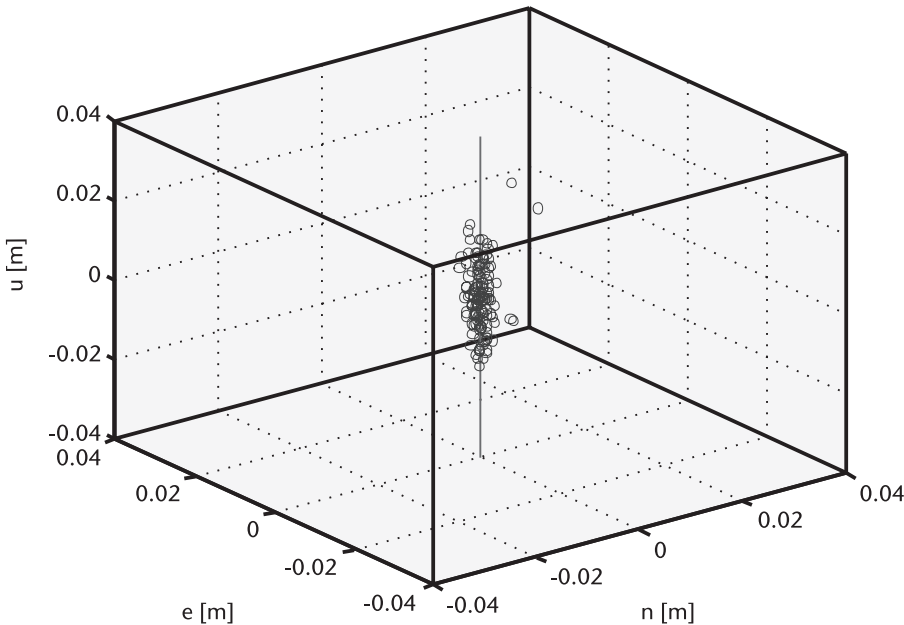


Figure 14.5 Coordinate residuals (n,e,u) with respect to the known user position, 24-hour data set processed in 144 ten-minute sessions [27].

Table 14.4 AR Statistic Summary, 24-Hour Data Set Processed in 144 Ten-Minute Sessions [27]

	AR Success [%]	Mean Time-to-fix [seconds]	Mean Time-to-fix Std [seconds]	Validation failure [%]
Iteration 1	97	236.1	164.7	0
Iteration 2	100	30.3	2.6	0

no need to maintain lock on the satellite signals when moving from one ground point to another.

14.4.1.4 Stop-and-Go Technique

This is a true *kinematic* technique because the user's receiver continues to track satellites while it is in motion. It is known as the stop-and-go technique because the coordinates of the receiver are only of interest when it is stationary (the "stop" part), but the receiver continues to function while it is being moved (the "go" part) from one stationary setup to the next. There are in fact three stages to the field operation:

1. *The initial ambiguity resolution:* This is carried out before the stop-and-go survey commences. The resolution of the CPH ambiguities (Section 14.2.1) by the software can be carried out using any method, but in general it is one of the following: (1) use a conventional static (or rapid-static) GNSS survey determines the baseline from the reference receiver to the first of the ground points occupied by the user's receiver; or (2) set up both receivers over a known baseline, and derive the values of the CPH ambiguities in this way; or (3) employ a field procedure known as "antenna swap" allowing the software to resolve the CPH ambiguities over a very short (few meter length) baseline; or (4) resolve the ambiguities "on-the-fly"—OTF-AR (i.e., while the receiver is tracking satellites but the receiver/antenna is moving).
2. *The receiver in motion:* Once the ambiguities have been resolved, the survey task can begin. The user's receiver is moved from point to point, collecting just a minute or so of data. It is vital that the antenna continues to track the satellites: in this way the resolved ambiguities are valid for all future *unambiguous* phase-range observations. However, if a cycle slip does occur, then the CPH ambiguities have to be reinitialized—bringing the receiver back to the last surveyed point, and redetermining the ambiguities can most easily be done using the "known baseline" method—see (1) above.
3. *The stationary receiver:* The unambiguous phase-range data is then processed in the DD mode to determine the coordinates of the user receiver relative to the static reference receiver (Section 14.2.1). *The trajectory of the antenna is not of interest, only the stationary ground points which are visited by the receiver.*

The technique is well-suited when many points close together have to be surveyed, and the terrain poses no significant problems in terms of signal disruption. The accuracy attainable is at the centimeter-level. The technique can also be implemented in real-time, as a variation of the GNSS-RTK field procedure.

14.4.1.5 Kinematic GNSS Technique

This is a generalization of the “stop-and-go” technique. Instead of only coordinating the stationary points and disregarding the trajectory of the antenna as it moves from point-to-point, the intention of *kinematic* surveying is to determine the position of the antenna *while it is in motion*. In other respects the technique is similar to the “stop-and-go” technique. That is, the CPH ambiguities must be resolved *before* starting the survey task, and the ambiguities must be reinitialized *during* the survey if a cycle slip occurs. However, for many applications such as the positioning of an aircraft or a ship, it is impractical to reinitialize the ambiguities if the moving antenna has to return to a stationary ground point. These days, the kinematic GNSS surveying technique routinely uses OTF-AR, making kinematic surveying techniques ideal for road centerline surveys, topographic surveys, hydrographic surveys, airborne applications, and many more. In addition, the widespread use of GNSS-RTK ensures that all coordinate information, for stationary ground points as well as antenna trajectories, is available without delay and can be used, as mentioned earlier, for time-critical applications.

14.4.2 GNSS Geodesy

This section presents examples of geodetic applications of GNSS and discusses operational modes of geodesy and surveying applications. In addition, examples of online GPS processing tools are presented together with example data sets.

14.4.2.1 Application: Studying the 2008 Sichuan Earthquake

The following is taken with permission from an article published by the Caltech Tectonics Observatory (<http://www.tectonics.caltech.edu/outreach/highlights/2008MayChinaEQ/>). Please see the on-line article for more detailed analysis. A summary is given below.

On May 12, 2008, a devastating earthquake struck China’s Sichuan province. The earthquake occurred in an area that is deforming as a result of the collision between two tectonics plates, the Indian plate and the Eurasian plate. This collision, which has been going on for 50 million years, is the cause of the high mountains and widespread seismicity observed throughout central Asia. It is this strain that led to the Sichuan quake. India has been moving northward at a rate of about 4 cm/year (2 inches/yr), which is about as fast as fingernails grow, pushing into central Asia and thus pushing Tibet. The 2008 Sichuan earthquake occurred in the eastern part of Tibet, indicated by the black star in Figure 14.6.

The motion of the land masses is shown in Figure 14.6. This velocity data was obtained from GPS stations located in the region. These stations enable us to measure surface velocity to within a fraction of a millimeter per year. The relatively fast northerly motion of India is evident, as is the somewhat slower easterly motion of Tibet.

A closer look at this region is shown in Figure 14.7. The rupture of the fault started in the mountains northwest of the city of Chengdu (indicated as a small white star) and then, over the next 50 seconds, traveled at least 200 km (100 miles) toward the northeast, tearing apart the land along the front of the mountain range.

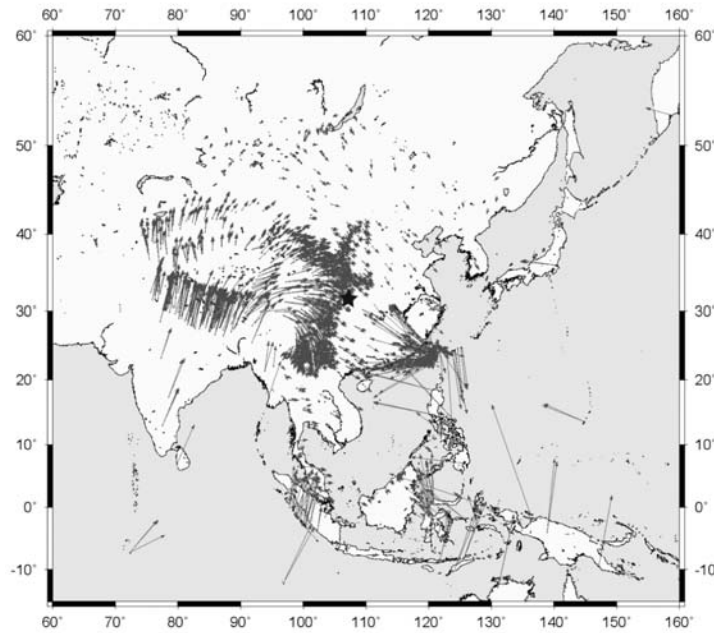


Figure 14.6 Velocity data from GPS stations located in the region. The black star indicates the location of the 2008 Sichuan earthquake. (Used with permission of Caltech Tectonics Observatory.)

In Figure 14.7, the white arrows show the horizontal motion of GPS stations located on the ground at those places, indicating the motion of the land. The arrows on the left-hand side are longer than those on the right-hand side, indicating that the land mass on the left is overtaking that on the right.

14.4.2.2 GNSS Geodesy Methodology

As already mentioned in Section 14.2.1, there is often a blurred boundary between GNSS surveying and GNSS geodesy. Nevertheless, some distinguishing characteristics of GNSS geodesy include the following:

- In the case of positioning applications, very high relative positioning accuracies are achieved—typically at the subcentimeter level—even though the interreceiver distances are in general far greater than is typical in GNSS surveying.
- These high-accuracy relative positioning applications include the determination of the coordinates of the reference stations themselves (e.g., for reference frame definition and maintenance); temporary stations established for the purpose of measuring ground deformation (e.g., due to earthquake or volcanic activity); tracking of buoys or other moving platforms (e.g., to address particularly challenging requirements); and the precise orbit determination (POD) of non-GNSS satellites (equipped with GNSS receivers).
- Positioning is not the only outcome of GNSS geodesy—other uses include POD of the GNSS satellite themselves; determination of Earth orientation parameters (EOPs), ionospheric (e.g., for “space weather” applications) and

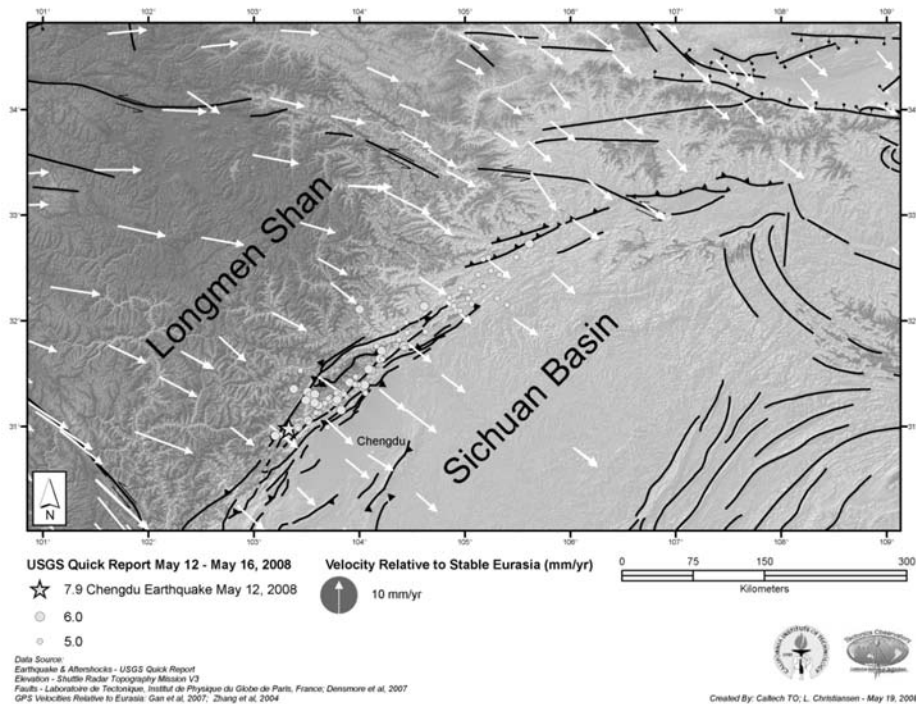


Figure 14.7 Map of fault region showing the epicenter of the first quake (star) as well as the location of aftershocks (circles) occurring within five days after the quake. (Used with permission of Caltech Tectonics Observatory.)

tropospheric (e.g., for “GNSS meteorology”) parameters from ground-based as well as satellite-borne GNSS receivers; precise time transfer; and a new technique for measuring ocean surface roughness (using bistatically reflected GNSS signals, see Chapter 16).

- The typical GNSS geodesy scenario is based on: (1) static positioning, (2) with long, continuous GNSS datasets (6–24 hrs is typical), (3) processed after data download (i.e., not in real-time), (4) using sophisticated scientific software packages, (5) that combine data from many simultaneously operating GNSS receivers in a rigorous network adjustment. (There are variations from such a scenario, including kinematic positioning of moving receivers, but these are not as common as the static positioning mode.)
- Much more complex mathematical CPH observable models are implemented inside current GNSS geodesy data processing software. The software typically accepts a priori information on satellite orbits, and perhaps other model parameters such as EOPs, reference station coordinates, tropospheric and ionospheric delays, obtained from the IGS. Furthermore, such software also permit the estimation of additional parameters—depending upon the applications and scenarios mentioned above—such as for *residual* GNSS satellite orbit error, tropospheric delay biases, and other systematic error terms.

The typical static scenario referred to above, in which the objective is the determination of the coordinates of GNSS receivers, is similar to the very first civilian application of GPS. The difference is that the relative accuracies have improved by several orders of magnitude; the cost of GNSS hardware has dropped to approximately one tenth it was in the early 1980s; there is nowadays greater use of permanent CORS infrastructure; and the IGS provides high-accuracy geodetic products. Consequently GNSS geodesy is now a routinely used technology, that is increasingly accessible to a diverse range of GNSS users working on all the continents on Earth. The new nonpositioning applications are further evidence of the versatility of the GNSS geodesy *methodology*.

14.4.2.3 Online GPS Processing Services

An example of how the GNSS geodesy CORS infrastructure, the IGS products and sophisticated data processing software is in innovative ways helping users is the development of noncommercial, web-based GPS/GNSS data processing services. Users can upload their recorded GNSS data to one of the “Web engine” servers. The user’s data may be: (1) combined with the data from CORS surrounding the user receiver’s location and processed in network mode with scientific software using IGS precise GNSS satellite orbit information, or (2) processed directly in precise point positioning (PPP) mode with special software that uses IGS precise GNSS satellite orbit and clock information. After processing the user is sent the final ITRF coordinates of the static receiver (and, in some cases, also the coordinates expressed in the national datum). The Web-based GPS/GNSS data processing services include JPL’s Auto-GIPSY (<http://milhouse.jpl.nasa.gov/ag/agfaq.html>), Scripps Coordinate Update Tool (SCOUT) (<http://sopac.ucsd.edu/cgi-bin/SCOUT.cgi>), Natural Resources Canada’s (NRC) on-line GPS Processing Service (http://www.geod.nrcan.gc.ca/products-produits/ppp_e.php), Geoscience Australia’s AUSPOS on-line GPS Processing Service (<http://www.ga.gov.au/geodesy/sgc/wwwgps/>), NGS OPUS and OPUS-RS (http://www.ngs.noaa.gov/OPUS/What_is_OPUS-RS.html; <http://www.ngs.noaa.gov/OPUS/>).

14.4.2.4 Example GPS Data Included on the DVD

Included on the DVD is a test dataset, including the RINEX observation data files; log0229n.08O (Observation) and log0229n.08N (Navigation), which can be processed using OPUS (static solution), OPUS-RS (rapid static solution) and the NRC GPS Processing Service (NRC GPS) (PPP solution). Also included is a summary document *readme_results.doc* containing the summary outputs of the OPUS, OPUS-RS and NRC GPS solutions, and the directories *PPP_ITRF05* and *PPP_NAD83* contain the NRC GPS detail processing reports.

Notice that for OPUS-RS the enclosed RINEX data file must be truncated to 30 minutes, and a blank line must be left in the end of the RINEX data file. In addition, the TIME OF FIRST OBS and the TIME OF LAST OBS in the RINEX header file must be changed accordingly. The first 30 minutes of the original data file has been clipped to the file log0229n_short.08O for processing using OPUS-RS.

14.5 The Future: The Next-Generation GNSS

More satellites and more signals will be welcome by many user communities, especially surveying and geodesy. However, the challenge will be to integrate all these satellite systems into one *GNSS system of systems*. *Compatibility* and *interoperability* of the different GNSS, combined with the current and proposed augmentation systems (see Chapter 4), are key concerns. The following sections include speculative remarks from the point of view of surveying and geodesy applications on the improvements that could be expected as GNSS systems continue to improve.

14.5.1 The Benefits of More Satellites and Signals

GPS and GLONASS in combination have already demonstrated the benefits of extra satellites, and adding GALILEO and COMPASS satellites holds the potential for additional enhancements. In addition, if SBAS and Radio Navigation Satellite Service (RNSS) satellites also transmit interoperable signals, the overall system performance will increase even further. The benefits of the extra satellites and their signals can be categorized in terms of continuity, accuracy, efficiency, availability, and reliability [29].

14.5.1.1 Improved Continuity

Extra satellites improve *continuity*; GPS, GLONASS, GALILEO, and COMPASS being independent GNSS means major system problems—unlikely as they are—have a very remote possibility of occurring simultaneously. GNSS/RNSS signals are vulnerable to interference or jamming—hence positioning, navigation, and timing (PNT) capability will be almost guaranteed if there are enough satellites and signals with different frequencies and modulated code characteristics.

14.5.1.2 Improved Accuracy

Extra satellites and signals can improve *accuracy*. In fact, more satellites to observe means a given level of accuracy can be achieved sooner using more measurements in the receiver's positioning algorithm (see simulation example in Chapter 3). Also, position accuracy is less susceptible to the influence of satellite geometry, as PDOP everywhere will be very low and largely constant. The effects of multipath and interference/jamming could be mitigated through implementation of RAIM algorithms—ensuring that only measurements of high quality are processed.

14.5.1.3 Improved Efficiency of Positioning

Extra satellites and signals can improve *efficiency*; In the case of CPH-based positioning, the extra satellites and signals will significantly reduce the time required to resolve ambiguities. Simulation studies conducted by researchers tend to fall into two groups:

1. In one category of studies, it is the *superior performance of OTF-AR for Triple Carrier Ambiguity Resolution (TCAR) or Multiple Carrier Ambiguity*

Resolution (MCAR) algorithms that is emphasized. In addition to the speed and reliability of OTF-AR, another significant advantage of triple-frequency systems is that the density of the CORS network needed to support DGNSS operations can be greatly reduced.

2. The other category of simulation studies identifies the *total number of satellites*, and the *favorable satellite-user geometry*, as a critical factor improving OTF-AR performance. That is, it is preferable to track two common frequencies on 20–30 visible satellites simultaneously, rather than three or more frequency signals on a subset of visible satellites. OTF-AR speed and reliability is indeed improved, but it is not at all clear from such studies what is the relationship between (say) PDOP and the CORS network density.

14.5.1.4 Improved Availability

Extra satellites and signals can enhance signal *availability* at a particular location, crucial for users wanting PNT solutions in areas that do not satisfy the *open-sky* conditions. The ideal conditions for high productivity GNSS CPH-based positioning typically are: (1) minimum of five (GPS) or six (GPS+GLONASS) visible satellites, (2) good receiver-satellite geometry (low PDOP values), (3) PR and CPH observations on both the L1 and L2 frequencies, and (4) low multipath disturbance. Hence improved availability can be interpreted as having more trackable multifrequency signals, such that in almost any outdoor tracking scenario there will be enough unobstructed satellites/signals for successful CPH-based positioning.

14.5.1.5 Improved Reliability

Extra satellites and signals can improve *reliability*, as with extra measurements the data redundancy is increased, which helps identify measurement outliers. The current L2 GPS measurements by geodetic-grade receivers are noisier and are less continuous than those that will be made on either of the new signals L2C or L5, which are expected to enhance dual-frequency operation in the future. Moreover, more signals means that service is not as easily denied due to interference or jamming of one frequency or set of signals.

14.5.2 Improvements to the GNSS Infrastructure

GNSS surveying and GNSS geodesy will both benefit significantly from the extra GNSS/RNSS satellites, and their many additional signals. However, the impact of next-generation GNSS will not only be by virtue of receiver hardware able to track more satellites and signals.

14.5.2.1 The CORS Infrastructure

Many high-accuracy users will continue to use the differential mode of positioning, taking advantage of permanent continuously operating reference stations. However, given the wide range of GNSS/RNSS frequencies and signals that could be tracked

by the CORS ground infrastructure for surveying and geodesy, commencing in approximately 2013, there may be a hierarchy of CORS networks [17]:

- Tier 1 will be the CORS equipped with receivers that can track all the broadcast GNSS/RNSS frequencies/signals, established for ultra-high-accuracy networks to support geoscientific research and global reference frame definition, including for the global IGS tracking network.
- Tier 2 will be the high-accuracy CORS networks equipped with receivers that can also track all the broadcast GNSS/RNSS frequencies/signals, operated by national geodetic agencies for the purpose of maintaining national geodetic datums.
- Tier 3 will be the CORS equipped with lower cost receivers that can track, for example, the L1+L5 GNSS/RNSS frequencies/signals, that densify the national CORS networks, operated by agencies such as state governments, or private companies providing commercial DGNSS/RTK services.

The implication is that the density of GNSS receivers within the CORS networks will be variable. The Tier 1 and Tier 2 CORS stations operating top-of-the-line, multifrequency “GNSS system of systems” receivers will likely be established with relatively long interreceiver separations, of the order of several hundred kilometers or more. On the other hand, the Tier 3 CORS stations operating the lower cost dual-frequency receivers, will likely be established at a variety of closer spacings, from just several kilometers apart (e.g., to support structural deformation monitoring, and single-base RTK applications), to perhaps up to several tens of kilometers (e.g., to support most DGNSS and network-RTK users). Furthermore, increasingly GNSS operations (for all tiers of CORS networks) will be in support of real-time applications. What is also far from clear is who will be responsible for generating, marketing and quality control of the GNSS-RTK services.

14.5.3 Applications and the Future

GNSS surveying applications are likely to grow in two respects:

The user base of real-time, high-accuracy GNSS-RTK will expand significantly as: (1) low-cost dual-frequency (L1+L5) receivers become available, (2) the CORS ground infrastructure grows, and (2) wireless communications link coverage expands. The automation of construction, agricultural and heavy transport machinery and vehicles will require very reliable, robust, continuous centimeter-level accuracy positioning—this will be provided by next-generation GNSS-RTK. GNSS surveying will therefore become an indispensable tool for surveying, engineering, precise navigation, real-time mapping and imaging, and machine automation and robotics.

In the case of GNSS geodesy, there will continue to be a blurring of the distinction between GNSS surveying and GNSS geodesy. For example, increasingly the latter will be carried out in real time (such as the real-time IGS Pilot Project [18]), and will share much of the CORS infrastructure with the former. The non-positioning applications such as real-time atmospheric monitoring, discussed in Chapter 15, are likely to become even more important than is currently the case.

In addition, GNSS geodesy techniques will become more operationalized, will be capable of greater accuracy and consistently high reliability, and will contribute significantly to the geosciences in the coming decades. In particular, the GNSS technology, implemented through a global framework such as provided by the IGS, will be perhaps the most important component of the planned *Global Geodetic Observing System* (GGOS). The GGOS will be a vital system for *global change studies*, supporting many terrestrial and space-borne observing platforms, by providing the following geodetic capabilities: (1) reliable millimeter-level accuracy positioning in GNSS networks several thousands of kilometers across, (2) precise orbit determination of Earth-orbiting satellites, (3) time transfer at the subnanosecond level of accuracy, (4) remote sensing of the Earth's troposphere and ionosphere, and (5) the estimation of ocean wave parameters from reflected GNSS signals (<http://www.ggos.org/>)

The future of GNSS is bright for surveying and geodesy. Both in respect to the current user communities, who will progressively see improved capabilities at lower cost and complexity; as well as new applications that will benefit from the increased robustness and availability of high-accuracy GNSS solutions.

References

- [1] Rizos, C., "Making Sense of the GPS Techniques." in *Manual of Geospatial Science and Technology*, Chapter 11, pp. 146–161, J. Bossler, et al (eds.), New York: Taylor & Francis Inc., 2002.
- [2] Bossler, J., et al (eds.), *Manual of Geospatial Science and Technology*, New York: Taylor & Francis Inc., 2002.
- [3] ITRF2005 (2008), International Terrestrial Reference Frame 2005 home page http://itrf.ensg.ign.fr/ITRF_solutions/2005/ITRF2005.php, accessed 17 April 2008.
- [4] Seeber, G., *Satellite Geodesy* (Second edition), Walter de Gruyter Berlin, New York, 2003.
- [5] Grejner-Brzezinska, D. A., and C. Rizos, "Satellite-Based Geodesy," McGraw-Hill *Yearbook of Science & Technology*, pp. 331–334, 2009, New York: The McGraw-Hill Companies, Inc.
- [6] Kouba, J., and P. Héroux, "Precise Point Positioning Using IGS Orbit and Clock Products," *GPS Solutions*, Vol. 5, No. 2, 2001, pp. 12–28.
- [7] Wielgosz, P., D. A. Grejner-Brzezinska, and I. Kashani, "High-Accuracy DGPS and Precise Point Positioning Based on Ohio CORS Network," *Navigation*, Vol. 52, No. 1, 2005, pp. 23–28.
- [8] Grejner-Brzezinska, D. A., "Integration of Global Positioning Systems (GPS) to Agricultural Geophysics," Chapter 8 in *Handbook of Agricultural Geophysics*, B. Allred, J. Daniels, R. Eshani (eds.), Boca Raton, London, New York, Washington D.C.: CRC Press, 2008.
- [9] Walsh D. M. A., and P. Daly, "Precise Positioning using GLONASS: Using GLONASS Carrier Phase Observables for Centimetre-Level Positioning," *GIM International*, No. 12, 1998, pp. 32–35.
- [10] Wang, J., M. P. Stewart, and M. Tsakiri, "Modelling GLONASS measurements for precise positioning," *Survey Review*, Vol. 36, No. 282, 2001, pp. 110–120.
- [11] Teunissen, P., "Integer least-squares," in *International Association of Geodesy Symposia*, pp. 69–80, F. Sanso (ed.), Vol. 127, Springer, 2004.
- [12] Dedes, G., and C. Rizos, "GPS Mathematical Models for Single Point and Baseline Positioning," in *Manual of Geospatial Science and Technology*, Chapter 9, pp. 114–126, J. Bossler et al (eds.), Taylor & Francis, Inc., 2002.

- [13] Misra, P., and P. Enge, P., *Global Positioning System: Signals, Measurements and Performance*, Lincoln, MA: Ganga-Jamuna Press, 2001. www.gpstextbook.com.
- [14] Rizos, C., "Alternatives to Current GPS-RTK Services and Some Implications for CORS Infrastructure and Operations," *GPS Solutions*, Vol. 11, No. 3, pp. 151–158. 2007.
- [15] IGS (2008), International GNSS Service home page <http://igsceb.jpl.nasa.gov>, or <http://www.igs.org>, accessed 17 April 2008.
- [16] Dow, J., R. Neilan, and C. Rizos, "The International GNSS Service (IGS): Preparations for the coming decade," in *Proc. 20th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation*, Fort Worth, Texas, 25–28 Sept. 2007, pp. 2136–2144.
- [17] Rizos, C., "Multi-Constellation GNSS/RNSS From the Perspective of High Accuracy Users in Australia," *Journal of Spatial Science*, 53(2), 29–63, 2008.
- [18] IGS-RT (2008), International GNSS Service, Real-Time Pilot Project home page <http://www.rtigs.net>, accessed 25 April 2008.
- [19] Vollath, U., et al, "Multi-Base RTK Positioning Using Virtual Reference Stations," in *Proc. ION GPS*, 19–22 Sept. 2000, Salt Lake City, UT, pp. 123–131.
- [20] Rizos, C., "Network RTK Research and Implementation: A Geodetic Perspective," *Journal of Global Positioning Systems*, Vol. 1, No. 2, 2002, pp. 144–150.
- [21] Wanninger, L., "Virtual Reference Stations for Centimeter-Level Kinematic Positioning," in *Proc. ION GPS*, 24–27 Sept. 2002, Portland, OR, pp. 1400–1407.
- [22] Grejner-Brzezinska, D. A., I. Kashani, and P. Wielgosz, "On Accuracy and Reliability of Instantaneous Network RTK as a Function of Network Geometry, Station Separation, and Data Processing Strategy," *GPS Solutions*. Vol. 9, No. 3, 2005, pp. 179–193.
- [23] Kashani, I., D. A. Grejner-Brzezinska, and P. Wielgosz, "Towards Instantaneous Network-Based RTK GPS Over 100 km Distance," *Navigation*. Vol. 52, No. 4, 2005, pp. 239–245.
- [24] Langley R. B., "RTK GPS," *GPS World*, September 1998, pp. 70–74.
- [25] Blewitt, G., "A new tool for dense monitoring of crustal strain: GPS rapid static surveying," *Eos. Trans. Am. Geophys. U.*, Vol. 71, No. 17, 1990, p. 483.
- [26] Schwarz, C. R., "Heuristic weighting and data conditioning in the National Geodetic Survey rapid static GPS (RSGPS) software." *J. Surv. Eng.*, Vol. 134, No. 3, 2008, pp. 76–82.
- [27] Kashani, I., Wielgosz, P., Grejner-Brzezinska, D. A., and G. L. Mader, "A New Network-Based Rapid-Static Module for the NGS On-line Positioning User Service—OPUS-RS," in *Proc. of the ION Annual Meeting*, 27–29 June 2005, Cambridge, MA, pp. 928–936.
- [28] Grejner-Brzezinska, D. A., Wielgosz, P., Kashani, I., Mader, G. L., Smith, D. A., Robertson, D., and A. Komjathy, "Performance Assessment of the New Rapid-Static Module of the On-line Positioning User Service—OPUS-RS," in *Proc. ION GNSS*, 13–16 Sept. 2005, Long Beach, CA, pp. 2595–2605.
- [29] Rizos, C., M. Higgins, and S. Hewitson, "New GNSS Developments and Their Impact on Survey Service Providers and Users," *Spatial Sciences Conference*, Melbourne, Australia, 12–16 Sept. 2005, pp. 1100–1113, also available from http://www.gmat.unsw.edu.au/snap/publications/rizos_et al2005a.pdf, accessed 18 April 2008.
- [30] GLONASS Interface Control Document (ICD), Version 5.0, Moscow, 2002, <http://www.glonass-ianc.rsa.ru>.

Atmospheric Sensing Using GNSS Occultations

Chi O. Ao

15.1 Introduction

A GNSS radio occultation (RO) occurs when a receiver on-board a LEO spacecraft tracks a GNSS satellite as it sets or rises through the Earth's atmosphere (Figure 15.1). The recorded occulted signal phase and amplitude of the tracked satellite can then be analyzed to derive atmospheric refractivity, density, pressure, temperature, and humidity measurements. The RO technique was first used to probe planetary atmospheres starting in the 1960s [1, 2]. The success of these experiments led to an early study that investigated the possibility of applying RO measurements to the remote sensing of the Earth's atmosphere [3]. However, the need to launch both receivers and transmitters in orbits made this a costly proposition. The idea laid dormant until the middle of 1980s when the availability of dual-frequency radio signals from the GPS constellation made this concept very appealing both from a cost-effectiveness and scientific point of view [4, 5]. These initial studies,

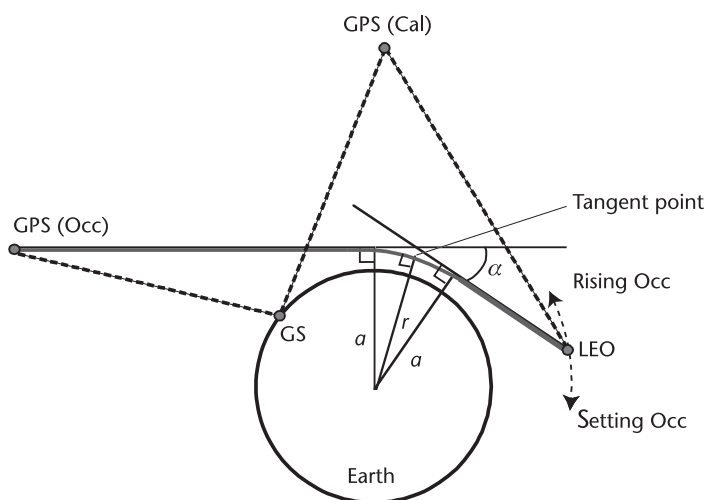


Figure 15.1 Geometry of GPS radio occultation. The signal transmitted by the GPS is bent, by an angle of α , as it travels through the Earth's ionosphere and atmosphere to the LEO. Under spherical symmetry, each ray can be identified uniquely by an impact parameter a .

along with more detailed analyses subsequently carried out [6, 7], have shown that GPS occultations possess several unique properties compared to other remote sensing instruments.

First, GPS RO measurements are self-calibrating. This means that there is no need to calibrate the measurements against “standards” that might drift over time, thus making GPS occultation an ideal climate dataset. Second, being an active limb-sounding instrument, GPS occultations yield atmospheric profiles with very high vertical resolution (<1 km). Third, the GPS frequencies were chosen by design so that they are not affected by the presence of clouds and precipitation. This ensures that the GPS RO measurements will not be degraded or biased in cloudy regions.

The first GPS occultation mission was the GPS/Meteorology (GPS/MET) experiment that operated in 1995–1997 [8]. It was equipped with the NASA/JPL Turbo-Rogue GPS receiver modified to acquire and track occultation signals. The demonstrable success of GPS/MET [9] led to the inclusion of “BlackJack” GPS RO receivers on the Challenging Minisatellite Payload (CHAMP) [10] and Satellite de Aplicaciones Cientificas-C (SAC-C) [11] satellites, which have been producing GPS RO data nearly continuously since 2001. The twin satellites of Gravity Recovery and Climate Experiment (GRACE) also carried the BlackJack receiver, and routine occultation measurements have been made since 2006. The most dramatic addition to the suite of GPS occultation sensors in space is the six-spacecraft Constellation Observing System for Meteorology, Ionospheric, and Climate (COSMIC), also known as FORMOSAT-3 constellation launched in April 2006 [12], which more than double the existing number of occultation soundings observed each day.

As we transition from the world of GPS to GNSS, an exciting future lies ahead for GPS occultation, with new GPS signals (L2C and L5) becoming available and the rapidly expanding number of transmitting sources from new navigation constellations such as Galileo and Beidou. Active developments of GNSS RO receiver technology carried out at JPL and elsewhere promise significant enhancements in the accuracy, precision, and throughput of the measurements.

This chapter provides a basic introduction of the GPS occultation technique, including: the characteristics of the measurements, the retrieval method, and a brief overview of the applications. Included at the end of the chapter is a description of the MATLAB/Octave scripts and GPS RO data from GRACE that can be used to perform simple example RO calculations. Readers are referred to the further reading section and references for more in-depth discussions of the topics introduced here. Finally, while the GPS RO technique is useful for remote sensing of both the ionosphere and the neutral atmosphere, the focus of this chapter will be on neutral atmosphere retrieval and applications.

15.2 Occultation Measurements

The configuration of a GPS occultation is shown in Figure 15.1. A receiver on-board a LEO satellite tracks one or more GPS signals as they pass through the Earth’s atmosphere. Due to the motion of the occulting GPS transmitters and LEO satellites,

the GPS signals essentially slice through different vertical layers of the atmosphere over the course of an occultation event. Over the period of this event, precise measurements of the signal carrier phase and amplitude as a function of time can be inverted to yield a vertical profile of the refractive index of the atmosphere.

The refractive index is directly related to electron density in the ionosphere, and is a function of pressure, temperature, and water vapor in the neutral atmosphere. Since the refractive index n is close to one, it is convenient to speak of the refractivity $N = (n - 1) \times 10^6$. The refractivity is approximately given by

$$N = a_1 \frac{P}{T} + a_2 \frac{P_{wv}}{T^2} - b \frac{n_e}{f^2} \quad (15.1)$$

where $a_1 = 77.6$, $a_2 = 3.73 \times 10^5$, $b = 40.3 \times 10^6$, T is the temperature in K, P is the pressure in mb, P_{wv} is the water vapor pressure in mb, and n_e is the electron density in $1/\text{m}^3$. The first two terms give the refractivity in the neutral atmosphere (also referred to as the “dry” and “wet” terms) [13], while the third term gives the leading contribution due to electron density in the ionosphere. The fundamental quantity measured in a GPS occultation is the time delay of the GPS signal, which varies as a function of the index of refraction of the intervening medium. The total measured delay (in units of length) can be written as [14]

$$L = -\frac{c}{f} \phi = r + \gamma + C_t + C_r + \nu \quad (15.2)$$

where c is the speed of light in vacuum, f is the carrier frequency for GNSS signal (such as the L1 or L2 GPS signals) in Hz, ϕ is the recorded phase in cycles, r refers to the range between the transmitter and receiver in vacuum, γ contains the extra delay due to the Earth’s ionosphere and atmosphere, C_t and C_r are the transmitter and receiver clock errors respectively, and ν refers to measurement noise that includes thermal noise and local multipath. This representation is conceptually the same as (3.1), but modified and rearranged slightly for this discussion.

The first step in data processing is to extract the phase delay due only to the ionosphere and atmosphere (known as the “excess phase delay”). This involves the removal of the terms r , C_t , and C_r , which requires precise knowledge of the GPS and LEO orbits and the clock errors of both the transmitter and the receiver. To “calibrate” the GPS and LEO clocks, the traditional method is to employ the so-called double-differencing technique where measurements from additional radio links are utilized, as indicated by the dashed lines in Figure 15.1. This can be achieved by forming the linear combination of

$$L_D = (L - L_{CL}) - (L_{CG} - L_{OG}) \quad (15.3)$$

where subscripts CL indicates the link between the calibrating GPS and the LEO, CG between the calibrating GPS and ground station, OG between the occulting GPS and ground station. This linear combination enables the cancellation of the C_t and C_r terms in the delay measurement. The disadvantage is that the measurement now contains additional noise from the calibrating links.

Since the deactivation of selective availability on May 2, 2000, the GPS clocks can be determined with sufficient accuracy that it becomes possible to obtain accurate retrievals by using the single-differencing technique, whereby the following link combination is computed [15]

$$L_S = L - L_{CL} \quad (15.4)$$

If the LEO clock has sufficient stability, such is the case for the receivers on board the GRACE satellites equipped with ultra stable oscillators (USOs), it is feasible to forgo any differencing altogether [16]. With single and zero differencing, the level of random noise in the delay measurements should be lower. The measurements are also not subject to systematic errors from the other calibrating links; however, the disadvantage is that the measurements might contain larger residual clock errors.

Figure 15.2 gives an example of GPS occultation measurements, as observed by one of the COSMIC satellites on December 31, 2007. This example is a typical setting occultation, with the ray tangent points located over the tropical south Pacific Ocean. The entire occultation lasted about two minutes, with a starting tangent height of about 125 km above the Earth's surface. The figure shows the 50-Hz one-second SNR (essentially the “amplitude” of the signal) and the double-differenced excess phase delay $\tilde{L} = L_D - r$ as a function of received time. It also shows the excess Doppler frequency measurement, $\Delta f = -(1/\lambda)d\tilde{L}/dt$, which is described in more detail in the Section 15.3.1. It is instructive to examine the time variation of the measurements in the course of the occultation. In the beginning of the occultation, the ray tangent point passes through the ionosphere and the upper stratosphere, where the medium is very tenuous. The amplitude of the signal is relatively unchanged, and the excess phase delay and Doppler are close to zero. As the ray descends the atmosphere, the stronger vertical gradient of the refractive index causes the ray to bend more, leading to an increase in the magnitude of excess Doppler as well as a decrease in amplitude (the latter phenomenon is known as “defocusing”). In the lower troposphere, as the amount of water vapor increases, fine-scale structures in the refractive index cause the signal to undergo strong defocusing and rapid amplitude and phase oscillations. The complex signal structures associated with the moist lower troposphere present rather challenging conditions for signal tracking as well as for retrievals, as explained below.

15.3 Atmospheric Retrievals

We now discuss how the occultation measurements can be used to derive atmospheric profiles. The retrieval process involves two steps. First, the excess phase delay and SNR measurements are used to calculate the bending angle profile $\alpha(a)$, where a is the ray impact parameter (Figure 15.1). Second, the bending angle profile is inverted to give the refractivity profile and subsequently the temperature, pressure, and humidity profiles.

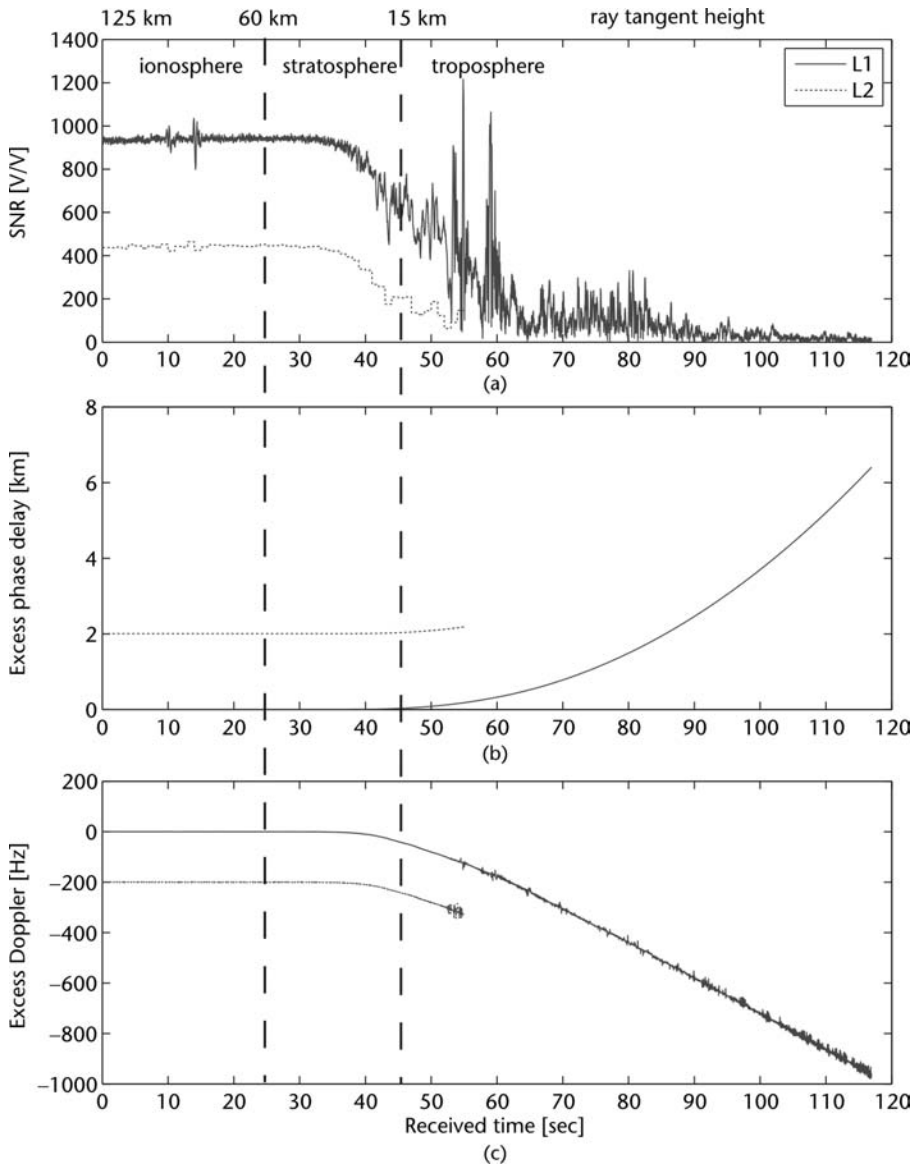


Figure 15.2 An example of GPS occultation measurements: (a) 1-sec SNR; (b) excess phase delay in kilometers; and (c) excess Doppler in Hertz. The L2 excess phase and Doppler have been shifted by 2 km and -200 Hz, respectively. Note that the L2 SNR is much lower than L1 due to the need for codeless tracking. Typically no useful L2 data are available in the lower and middle troposphere.

15.3.1 Derivation of Bending Angle Profiles

A key characterization of each occultation is the bending angle profile, which gives a geometric optics representation of the GPS signal propagation. In the traditional approach, this can be derived using the excess Doppler frequency [1, 14]

$$\lambda \Delta f = \mathbf{v}_t \cdot \hat{\mathbf{k}}_t - \mathbf{v}_r \cdot \hat{\mathbf{k}}_r - (\mathbf{v}_t - \mathbf{v}_r) \cdot \hat{\mathbf{k}} \quad (15.5)$$

where \mathbf{v}_t and \mathbf{v}_r are the velocities of the transmitter (t) and receiver (r), $\hat{\mathbf{k}}_t$ and $\hat{\mathbf{k}}_r$ are the unit vectors corresponding to the ray direction for the transmitter and receiver, and $\hat{\mathbf{k}}$ is the unit vector from the transmitter to the receiver (Figure 15.3).

Since $\hat{\mathbf{k}}_t$ and $\hat{\mathbf{k}}_r$ are unknown, there would normally be four independent parameters that need to be solved for (azimuth and elevation angles describing each direction vector). However, if we assume that the atmosphere is spherically symmetric, the ray is confined to a plane so that there are only two unknowns, and we can write (15.5) as

$$\lambda \Delta f = v_t \cos(\phi_t - \delta_t) - v_r \cos(\phi_r - \delta_r) - (v_t \cos \phi_t - v_r \cos \phi_r) \quad (15.6)$$

The unknowns angles δ_t and δ_r can be solved with the use of the Bouguer's formula (Snell's law in a spherically symmetric medium) [17]:

$$a = n(r_t) r_t \sin(\theta_t + \delta_t) = n(r_r) r_r \sin(\theta_r + \delta_r) \quad (15.7)$$

where a is the impact parameter (which is an invariant property of the ray in spherically symmetric medium, see Figure 15.1), $n(r_t)$ and $n(r_r)$ are the indices of refraction at the transmitter and receiver, respectively, which can be set to 1 to good approximation.

Equations (15.6) and (15.7) can be used to solve for the angles δ_t and δ_r , from which the ray bending α can be obtained

$$\alpha = \delta_t + \delta_r \quad (15.8)$$

At high altitudes, the atmosphere is tenuous so that the angles δ_t and δ_r are small. It is convenient to start solving for these angles from the top down in an iterative fashion [1]. To this end, we rewrite the unknowns as $\delta_t = \tilde{\delta}_t + \epsilon_t$ and $\delta_r = \tilde{\delta}_r + \epsilon_r$, where ϵ_t and ϵ_r are small. Using these definitions in (15.6) and (15.7) and keeping only the leading orders of ϵ_t and ϵ_r , we obtain a set of linear equations, which can be solved to yield

$$\epsilon_t = \frac{b_2 A - a_2 B}{a_1 b_2 - a_2 b_1} \quad (15.9)$$

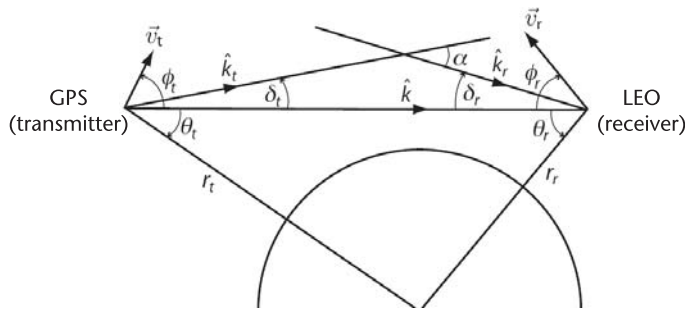


Figure 15.3 Geometry for deriving the bending angle and impact parameter from Doppler shift. The unknowns angles δ_t and δ_r can be solved with the use of the Bouguer's formula (Snell's law in a spherically symmetric medium) [17].

$$\epsilon_r = \frac{b_1 A - a_1 B}{a_1 b_2 - a_2 b_1} \quad (15.10)$$

where $A = \lambda \Delta f - v_t [\cos(\phi_t - \tilde{\delta}_t) - \cos \phi_t] + v_r [\cos(\phi_r - \tilde{\delta}_r) - \cos \phi_r]$, $B = r_r \sin(\theta_r + \tilde{\delta}_r) - r_t \sin(\theta_t + \tilde{\delta}_t)$, $a_1 = v_t \sin(\phi_t - \tilde{\delta}_t)$, $a_2 = v_r \sin(\phi_r - \tilde{\delta}_r)$, $b_1 = r_t \cos(\theta_t + \tilde{\delta}_t)$, and $b_2 = r_r \cos(\theta_r + \tilde{\delta}_r)$. Starting with initial guesses for $\tilde{\delta}_r$ and $\tilde{\delta}_t$ (say using the values at the layer just above it and zero at the very top layer), it is simple to solve for the corrections ϵ_t and ϵ_r iteratively until the desired precision is reached.

The traditional procedure for computing bending angles from excess Doppler alone is very effective except when the signal passes through the moist lower troposphere (below 6 km or so in the tropics and mid-latitudes). This approach, which implicitly assumes that only a single ray arrives at the receiver at each time, breaks down when rays with different travel paths converge at the receiver due to the presence of atmospheric structures with fine vertical scales. To untangle the multiple rays, it is necessary to transform the measurements from the time domain to another domain where the rays are uniquely identified. Two methods have been proposed to address this problem: the canonical transform (CT) method [18] and the full spectrum inversion (FSI) method [19]. The validity of these methods relies on the fact that under spherically symmetric conditions, the impact parameter a uniquely identifies each ray. Unlike the traditional approach, these new methods require the use of the SNR/amplitude data as well as the phase data. The complex wave field $A \exp(i\phi)$ is transformed via a Fourier integral operator to another representation (impact parameter for CT, Doppler frequency for FSI), where the derivative of the unwrapped transformed phase essentially yields the bending angle.

Another limitation of the traditional method is that its vertical resolution is limited by Fresnel diffraction. For typical GPS occultation geometry, the Fresnel diameter is about 1 km in the upper troposphere and lower stratosphere region. With CT or FSI, it is possible to achieve theoretically vertical resolution better than 100m.

15.3.2 Ionospheric Calibration

The bending angle profiles from GPS L1 and L2 signals differ due to the dispersive nature of the ionosphere. Since our goal is to derive neutral atmospheric profiles, it is important to remove the contribution from the ionosphere. The difference between L1 and L2 bending angle profiles provides an effective way to remove the dominant effects of the ionosphere from the neutral atmosphere. The leading order of ionospheric bending (which is proportional to $1/f^2$) can be removed by forming the following linear combination [20]

$$\alpha_n(a) = \left[\frac{f_1^2}{f_1^2 - f_2^2} \right] \alpha_1(a) - \left[\frac{f_2^2}{f_1^2 - f_2^2} \right] \alpha_2(a) \quad (15.11)$$

However, due to the fact that the L2 SNR is much lower than L1 (Figure 15.2), it is desirable to smooth the L2 data with a larger time interval. Thus it is more preferable to modify the above expression as [14]

$$\alpha_n(a) = \alpha_1(a) + \left[\frac{f_2^2}{f_1^2 - f_2^2} \right] (\bar{\alpha}_1(a) - \bar{\alpha}_2(a)) \quad (15.12)$$

where $\bar{\alpha}_1$ and $\bar{\alpha}_2$ denotes the L1 and L2 bending angles smoothed over larger time intervals.

Figure 15.4(a) shows the derived bending angle profiles for the example occultation of Figure 15.2. The L1 bending angle profile is obtained using the CT method. In addition, the L2 bending angle is obtained via the standard method and smoothed over two-second intervals (approximately 4 km). The neutral atmosphere bending angle obtained after ionospheric correction is also shown. The ionospheric calibration is especially critical for obtaining accurate retrievals in high altitudes. Notice that the neutral bending angles become very small and susceptible to measurement noise and residual ionospheric error above ~50 km.

15.3.3 Derivation of Atmospheric Profiles

Under the assumption of spherically symmetric atmosphere, the bending angle is related to the refractive index as

$$\alpha(a) = -2a \int_a^\infty \frac{1}{\sqrt{a'^2 - a^2}} \frac{d \ln n}{da'} da' \quad (15.13)$$

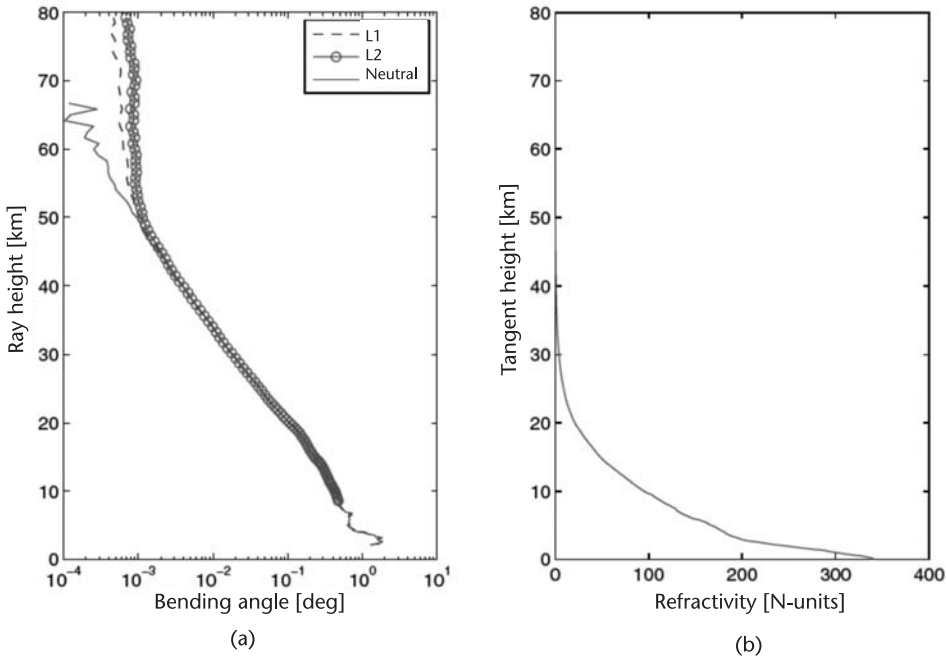


Figure 15.4 (a) Bending angle profiles for L1 and L2 measurements and the neutral bending angle computed according to (15.12) and (b) refractivity profile derived via the Abel inversion of the neutral bending angle.

This can be inverted to give the refractive index by using the Abel transform

$$\ln n = \frac{1}{\pi} \int_a^{\infty} \frac{\alpha(a')}{\sqrt{a'^2 - a^2}} da' \quad (15.14)$$

In practice, the bending angle profile is available only below a certain height a_u (mainly limited by measurement noise), so that the bending angle profile above a_u needs to be modeled. A popular solution for this problem is to replace the noisy bending angles at high altitudes (typically above 40–60 km) with “modeled” bending angles obtained from a climatology such as MSIS. The influence of the modeled bending angles decreases as the altitude decreases. The possible problem with this approach is that the retrievals could become biased towards the adopted climatology. An alternative, climatology-independent approach is to extrapolate the data at lower altitudes upward to altitudes where the data are not trustworthy. The problem with the extrapolation approach is that it relies on questionable assumptions regarding the characteristics of the stratosphere and mesosphere. Instead of selecting a fixed height a_u to switch from measurement to model, an effective strategy is to use a linear combination of the measurement and the model, with coefficients chosen in a statistically optimized manner according to the noise property of the bending angles [21]. Figure 15.4(b) illustrates the refractivity profile computed from the neutral bending angle using (15.14). Here, the bending angle above 50 km is modeled by using exponential extrapolation of the data between 40 and 50 km.

Refractivity is related to the temperature, pressure, and water vapor pressure through (15.1). Thus refractivity alone cannot determine these geophysical quantities without additional constraints. In the upper troposphere and the stratosphere, the amount of water vapor is sufficiently small that its contribution to the refractivity is negligible. In this case, there are still two unknowns P and T . To solve for them, one can assume hydrostatic equilibrium, which gives

$$\frac{dP}{dz} = -\rho g \quad (15.15)$$

where ρ is the air density related to the temperature and pressure through the ideal gas law:

$$\rho = \frac{m_d}{R} \frac{P}{T} + \frac{(m_w - m_d)}{R} \frac{P_w}{T} \quad (15.16)$$

where m_d , m_w are the mean molecular masses of dry air and water vapor, respectively, and R is the universal gas constant. Using (15.1) and (15.16), we can recast (15.15) as

$$\frac{dP}{dz} = -\frac{m_d g}{a_1 R} N + \frac{a_2 m_d g}{a_1 R} \frac{P_{wv}}{T^2} + \frac{(m_d - m_w) g}{R} \frac{P_{wv}}{T} \quad (15.17)$$

The procedure for solving (15.17) is to start at the maximum height z_m , typically high in the stratosphere where the water vapor contribution is insignificant. Ignoring the second and third terms on the right-hand side of (15.17), we obtain

$$P(z) = P(z_m) + \frac{m_d}{a_1 R} \int_z^{z_m} dz' g(z') N(z') \quad (15.18)$$

from which the temperature can be determined

$$T(z) = a_1 \frac{P(z)}{N(z)} \quad (15.19)$$

This solution is commonly referred to as the “dry retrieval” in GPS occultation parlance. The dry retrieval is not accurate in the lower and middle troposphere when the wet term in the refractivity becomes nonnegligible. In this case, it is not possible to solve for the temperature and water vapor independently without further constraint.

One common solution is to assume that the temperature profile is known and solve for water vapor pressure iteratively via (15.17) and (15.1). This approach works well in practice because the weather analyses from European Centre for Medium-range Weather Forecasts (ECMWF) and National Center for Environmental Prediction (NCEP) have more accurate temperature than water vapor information. Figure 15.5 shows the derived “dry” temperature and water vapor profiles from the refractivity profile of Figure 15.4(b) derived in this fashion. Also shown for comparison are the colocated ECMWF temperature and water vapor profiles. The (dry) temperature retrieval is obtained by initializing the hydrostatic equation at $z_m = 40$ km with ECMWF. In general, it is useable down to about 240–250K in the lower troposphere (about 8 km in the tropics), when water vapor becomes more important. Below this height, ECMWF temperature is assumed, and the water vapor profile is derived from GPS refractivity.

Instead of the iterative method described above, an alternative approach is to employ the one-dimensional variational (1D-Var) retrieval method for the refractivity profile, which can provide an optimal (in the least-square sense) estimation of T and P_{uv} based on an a priori along with assumed error characteristics for the refractivity measurements and the a priori [22].

15.4 Weather and Climate Applications

GPS occultation measurements have proven to be very useful for both weather and climate applications. We highlight a few areas in this section.

A great deal of research has been carried out on the methodology and effectiveness of assimilating GPS RO data into global as well as regional numerical weather prediction (NWP) models. The simplest, least computationally expensive approaches are to assimilate the bending angle or refractivity profiles. Several

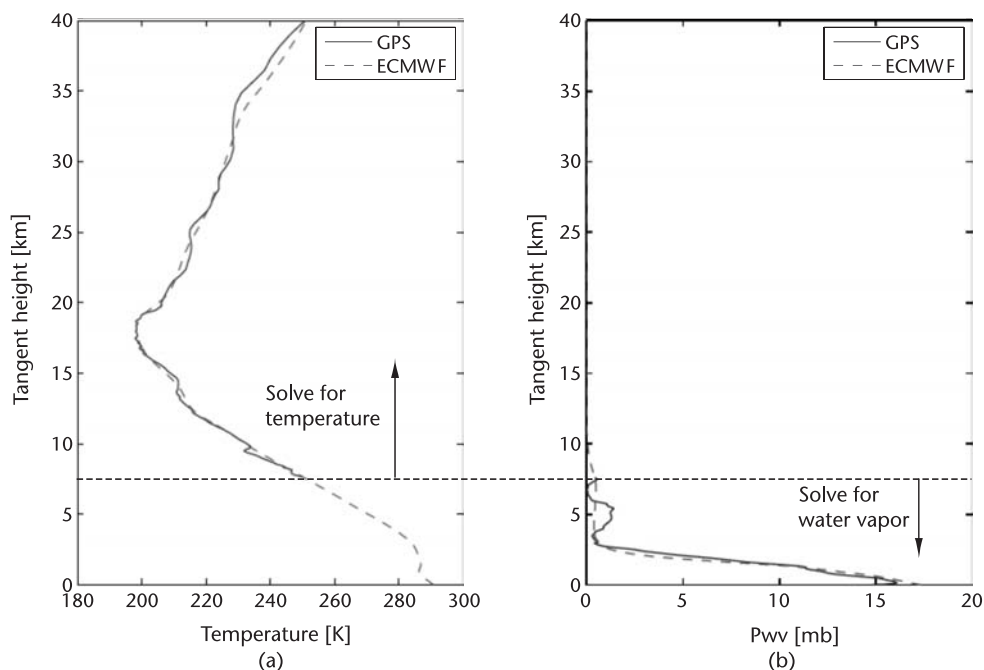


Figure 15.5 (a, b) Temperature and water vapor profiles derived from our example occultation. Also shown for comparison are the colocated ECMWF profiles. Note the much finer vertical structures revealed in the GPS occultation retrievals.

studies have demonstrated convincingly that assimilation of GPS data yielded positive impacts on the forecasts [23, 24]. A number of weather centers around the world, among them ECMWF and NCEP, have been ingesting GPS bending angle and/or refractivity measurements into their operational NWP models since 2006.

One of the most promising applications of GPS occultation is its potential for serving as a climate benchmark [25]. Detecting small long-term climate trends requires that the measurements are very precise and do not vary as a result of changes in instrumentation. The unique self-calibrating nature of the GPS occultation measurements means that the measurements are not subject to biases between different satellites or time-dependent drifts due to orbital decay—problems that have plagued the efforts to infer long-term temperature trends from the Microwave Sounding Unit (MSU) measurements. Analysis of temperature profiles from colocated CHAMP and SAC-C occultations found them to be consistent to 0.1K in the mean and 1K in standard deviation [11]. The high precision of the GPS measurements was also confirmed in a recent study [26] of colocated occultations from COSMIC.

Another scientific area where GPS occultation has proven useful is in delineating the characteristics of the tropopause, which separates the convectively mixed troposphere (where temperature decreases with height) and the convectively stable stratosphere (where temperature increases with height). The tropopause plays a crucial role in tropical dynamics and the vertical transport of trace gases; moreover, the tropopause height can be a sensitive indicator of climate change. The high vertical

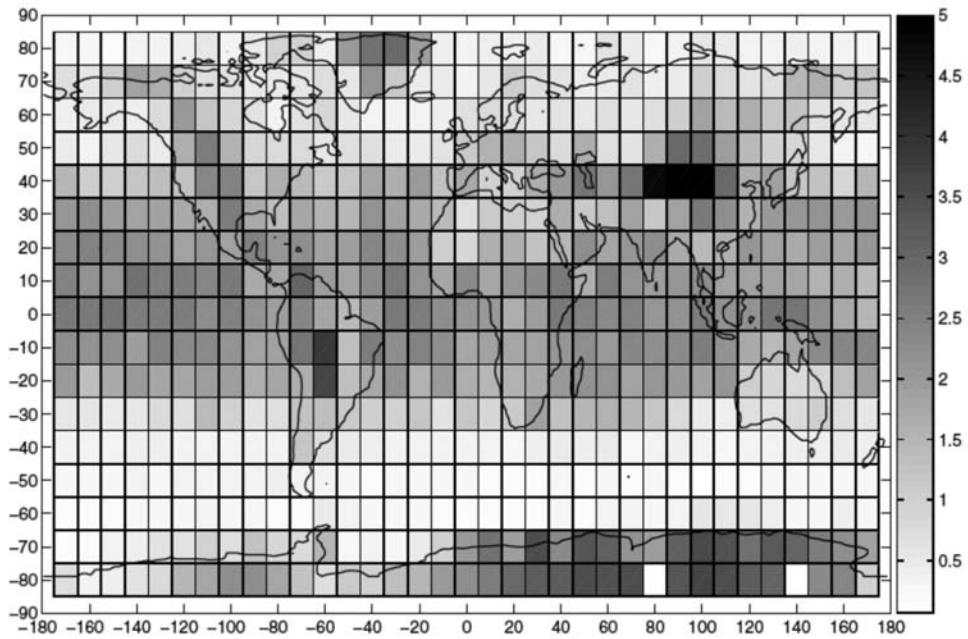
resolution that GPS occultation temperature profiles make them especially suitable for studying the tropopause [27–30].

15.5 Recent Advances

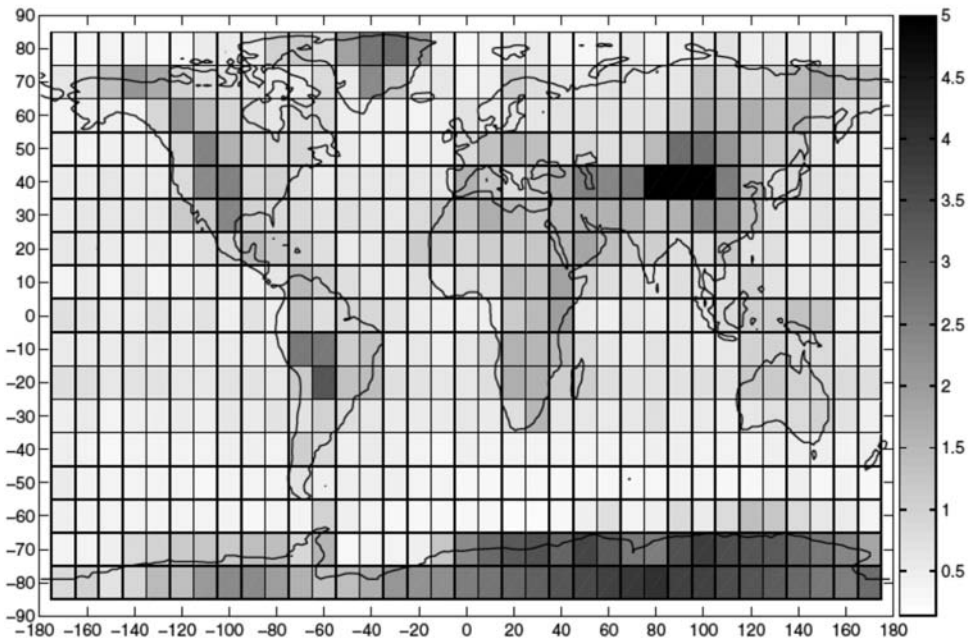
The past decade has seen a number of advances in the GPS RO technique, most of which were devised to deal with the very challenging conditions that exist for signals passing through the Earth's lower troposphere. Section 15.3.1 discusses the problem of atmospheric multipath and its solution through the methods of CT or FSI. A related, more ominous problem concerns the tracking of the signals through these regions. Traditionally, the modulated GPS signals are tracked using a PLL and DLL that relies on a feedback process to keep the reference signal sufficiently close (See Chapters 2 and 5). However, in the lower troposphere, large SNR drop-offs could be present for extended periods (Figure 15.2), leading to large errors in the estimation of the reference signal. These errors, in turn, cause significant number of cycle slips and amplitude suppression. It is believed such closed-loop (CL) tracking errors were the dominant source of the negative refractivity bias (N-bias) observed with GPS/MET data [9], and later in the CHAMP and SAC-C data [31, 32]. As a result of tracking errors, the profiles are often truncated above 3 km altitude.

Furthermore, for rising occultations, the signal is emerging from the lower troposphere. Therefore, the receiver would have to acquire and lock on to a very weak signal with high dynamics. By the time the receiver is able to acquire the signal via PLL, the tangent point has already risen to about 10 km, making the rising occultation not useful in sensing the lower-troposphere. This limitation is overcome by use of open-loop (OL) tracking, first implemented on SAC-C and later adapted on COSMIC [33]. OL tracking relies on a priori models of atmospheric Doppler and delay to track the occulting GPS signal. These so-called OL models must be sufficiently accurate for this process to work. At the 50-Hz sampling currently employed on SAC-C and COSMIC, the Doppler model should remain within 25 Hz of the true Doppler to avoid aliasing and the range delay model should be well within a GPS C/A code cycle (~300m) for OL tracking of the signal. Figure 15.6 shows the comparison between minimum profile heights from CL versus OL data from SAC-C. It is evident that OL dramatically improves the sampling in the lowest 4-km altitude, especially over the tropics. Over 80% of the SAC-C OL profiles now reach below 2 km altitude in the tropics, compared to only 50% achieved under CL tracking.

Another source of the observed negative N-bias is thought to be arising from a nonuniqueness problem affecting Abel inversions in the presence of elevated ducts or so-called superrefraction layers. This occurs when the vertical refractivity gradient dN/dz is less than about -157 N-units/km. Under such a condition, there exists an infinite number of refractivity profiles that give the same bending angle profile, and the Abel-inverted profile is always negatively biased below the duct [32, 34, 35]. A practical solution to this problem has not yet been found, although a promising approach has been proposed [36].



(a)



(b)

Figure 15.6 (a) Minimum profile heights in kilometers from SAC-C CL data from 2001–2003. The values correspond to the median values of minimum mean-sea-level heights from profiles within each $10^\circ \times 10^\circ$ latitude-longitude box. (b) Minimum profile heights from SAC-C OL data from 2006–2007.

Besides advances related to the lower troposphere, it is also worthwhile to point out the interesting work that has begun with modernization of the GPS constellation. The existence of new signals (L2C and L5) offers new possibilities in occultation data processing. Even though the GPS receivers on COSMIC were launched to track the L2P signal, it is possible to modify the receiver software to enable L2C open-loop tracking in the lower troposphere. A demonstration of this capability has been carried out, and preliminary analysis showed that the L2 data remained useful down to the surface [37]. The L2C data also offers stronger SNR at higher altitudes and should help enhance the accuracy of upper atmospheric retrievals.

15.6 Scripts and Data Included on the DVD

Here we will describe some simple MATLAB/Octave scripts that can be used to retrieve refractivity and temperature profiles from excess phase data. The included data were derived from GPS RO measurements from GRACE. Additional excess phase data can be downloaded from <http://genesis.jpl.nasa.gov>. GPS and GRACE orbits needed for the retrievals are also included on the DVD.

The primary scripts included are the following:

- `getBendingAngle.m` computes the bending angle profile $\alpha(a)$ given a time series of excess phase delays along with the GPS and LEO positions and velocities using (15.6)–(15.8).
- `getNeutralBending.m` takes the L1 and L2 bending angle profiles and removes the ionospheric contribution using (15.12).
- `getRefractivityProfile.m` computes the refractivity profile given the neutral bending angle profile using (15.14).
- `getDryTemperature.m` computes the dry temperature and pressure profiles from the refractivity profile using (15.18) and (15.19).

The file `gpsroDemo.m` illustrates the usage of these scripts by applying these scripts on a GRACE occultation. Figure 15.7 shows the output from this program.

15.7 Further Reading

The following references can be consulted for more in-depth discussion of the GPS RO technique. In particular, the reader is encouraged to study Kursinski et al (1997) for its comprehensive error analysis, an important topic that we did not cover in this chapter due to lack of space.

- Kursinski, E. R., et al, “Observing Earth’s Atmosphere with Radio Occultation Measurements Using the Global Positioning System,” *J. Geophys. Res.*, 102(D19):23429–23465, 1997.

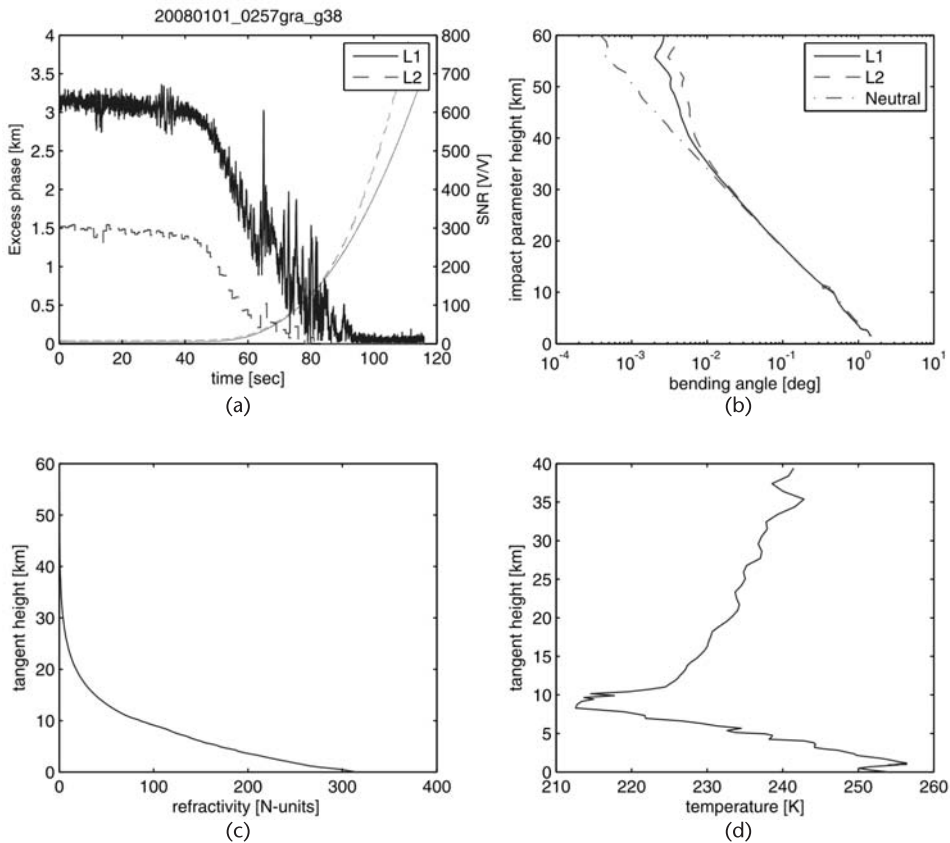


Figure 15.7 An example output from the scripts discussed in this section. (a) L1 and L2 phase delay data that serve as input to the scripts. (b) L1, L2, and neutral bending angle profiles computed by calling `getBendAngle` and `getNeutralBending`. (c) Refractivity profile obtained from Abel inversion integral by calling `getRefractivityProfile`. (d) Dry temperature profile derived from the refractivity profile by calling `getDryTemperature` and initializing the hydrostatic equation with temperature equal to 240K at 40 km.

- Hajj, G. A., et al, “A Technical Description of Atmospheric Sounding by GPS Occultation,” *J. Atmospheric and Solar-Terrestrial Phys.*, 64(4):451–469, 2002.
- Melbourne, W. G., *Radio Occultations Using Earth Satellites*, Hoboken, NJ: Wiley-Interscience, 2005.

References

- [1] Fjeldbo, G., A. J. Kliore, and V. R. Eshleman, “The Neutral Atmosphere of Venus as Studied with the Mariner V Radio Occultation Experiments,” *Astron. J.*, 76:123–140, 1971.
- [2] Tyler, G. L., “Radio Occultation Experiments in the Outer Solar System with Voyager,” *Proc. IEEE*, 75:1404–1431, 1987.
- [3] Lusignan, B., et al, “Sensing the Earth’s Atmosphere with Occultation Satellites,” *Proc. IEEE*, 57(4):458–467, April 1969.

- [4] Gurvich, A. S., and T. G. Krasil'nikova, "Navigation Satellites for Radio Sensing of the Earth's Atmosphere (in Russian)," *Soviet J. Remote Sensing*, 6:89–93, 1987.
- [5] Yunck, T. P., G. F. Lindal, and C. H. Liu, "The Role of GPS in Precise Earth Observation," *IEEE Position, Location and Navigation Symposium*, Orlando, FL, Nov. 29–Dec. 2, 1988.
- [6] Melbourne, W. G., et al, *The Application of Spaceborne GPS to Atmospheric Limb Sounding and Global Change Monitoring*, JPL Publication 94-18, 1994.
- [7] Kursinski, E. R., et al, "Observing Earth's Atmosphere with Radio Occultation Measurements Using the Global Positioning System," *J. Geophys. Res.*, 102(D19):23429–23465, 1997.
- [8] Ware, R., et al, "GPS Sounding of the Atmosphere from Low Earth Orbit—Preliminary Results," *Bull. Am. Meteorol. Soc.*, 77:19–40, 1996.
- [9] Rocken, C., et al, "Analysis and Validation of GPS/MET Data in the Neutral Atmosphere," *J. Geophys. Res.*, 102(D25):29849–29866, 1997.
- [10] Wickert, J., et al, "Atmospheric Sounding by GPS Radio Occultation: First Results from CHAMP," *Geophys. Res. Lett.*, 28(17):3263–3266, 2001.
- [11] Hajj, G. A., et al, "CHAMP and SAC-C Atmospheric Occultation Results and Intercomparisons," *J. Geophys. Res.*, 109:D06109, 2004.
- [12] Anthes, R. A., et al, "The COSMIC/FORMOSAT-3 Mission: Early Results," *Bull. Am. Meteorol. Soc.*, 89:313–333, 2008.
- [13] Smith, E. K., and S. Weintraub, "The Constants in the Equation for Atmospheric Refractive Index at Radio Frequencies," *Proc. IRE*, 41:1035–1037, 1953.
- [14] Hajj, G. A., et al, "A Technical Description of Atmospheric Sounding by GPS Occultation," *J. Atmospheric and Solar-Terrestrial Phys.*, 64(4):451–469, 2002.
- [15] Wickert, J., et al, "GPS Radio Occultation with CHAMP: Atmospheric Profiling Utilizing the Space-Based Single Difference Technique," *Geophys. Res. Lett.*, 29(8):1187, 2002.
- [16] Beyerle, G., et al, "GPS Radio Occultation with GRACE: Atmospheric Profiling Utilizing the Zero Differencing Technique," *Geophys. Res. Lett.*, 32:L13806, 2005.
- [17] Born, M., and E. Wolf, *Principles of Optics* (Seventh Edition), Cambridge, U.K.: Cambridge University Press, 1999.
- [18] Gorbunov, M. E., "Canonical Transform Method for Processing Radio Occultation Data in the Lower Troposphere," *Radio Sci.*, 37(5):1076, 2002.
- [19] Jensen, A. S., et al, "Full Spectrum Inversion of Radio Occultation Signals," *Radio Sci.*, 38(3):1040, 2003.
- [20] Vorob'ev, V. V., and T. G. Krasil'nikova, "Estimation of the Accuracy of the Atmospheric Refractive Index Recovery from Doppler Shift Measurements at Frequencies Used in the NAVSTAR System," *Phys. Atmos. Ocean*, 29:602–609, 1994.
- [21] Healy, S. B., "Smoothing Radio Occultation Bending Angles Above 40 km," *Ann. Geophys.*, 19:459–468, 2001.
- [22] Healy, S. B., and J. R. Eyre, "Retrieving Temperature, Water Vapour and Surface Pressure Information from Refractive-Index Profiles Derived by Radio Occultation: A Simulation Study," *Quart. J. Royal Met. Soc.*, 126(566):1661–1683, 2000.
- [23] Healy, S. B., and J. N. Thépaut, "Forecast Impact Experiment with GPS Radio Occultation Measurements," *Quart. J. Royal Met. Soc.*, 132:605–623, 2006.
- [24] Cucurull, L., et al, "Preliminary Impact Studies Using Global Positioning System Radio Occultation Profiles at NCEP," *Monthly Weather Rev.*, 136:1865–1877, 2008.
- [25] Goody, R., J. Anderson, and G. North, "Testing Climate Models: An Approach," *Bull. Am. Meteorol. Soc.*, 79(11):2541–2549, 1998.
- [26] Schreiner, W., et al, "Estimates of the Precision of GPS Radio Occultations from COSMIC/FORMOSAT-3," *Geophys. Res. Lett.*, 34:L04808, 2007.

- [27] Nishida, M., et al, "Seasonal and Longitudinal Variations in the Tropical Tropopause Observed with the GPS Occultation Technique (GPS/MET)," *J. Met. Soc. Japan*, 78:691–700, 2000.
- [28] Randel, W. J., F. Wu, and W. R. Rios, "Thermal Variability of the Tropical Tropopause Region Derived from GPS/MET Observations," *J. Geophys. Res.*, 108(D1):4024, 2003.
- [29] Schmidt, T., et al, "Tropical Tropopause Parameters Derived from GPS Radio Occultation Measurements with CHAMP," *J. Geophys. Res.*, 109:D13105, 2004.
- [30] Borsche, M., G. Kirchengast, and U. Foelsche, "Tropical Tropopause Climatology as Observed with Radio Occultation Measurements from CHAMP Compared to ecmwf and ncep Analyses," *Geophys. Res. Lett.*, 34:L03702, 2007.
- [31] Marquardt, C., et al, "Validation and Data Quality of CHAMP Radio Occultation Data," in *First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies*, ed. by C. Reigber, H. Lühr, and P. Schwintzer, Berlin: Springer-Verlag, 2003, pp. 384–396.
- [32] Ao, C. O., et al, "Lower Troposphere Refractivity Bias in GPS Occultation Retrievals," *J. Geophys. Res.*, 108(D18):4577, 2003.
- [33] Ao, C. O., et al, "Rising and Setting GPS Occultations by Use of Open Loop Tracking," *J. Geophys. Res.*, 114:D04101, 2009.
- [34] Sokolovskiy, S. V., "Effect of Super Refraction on Inversions of Radio Occultation Signals in the Lower Troposphere," *Radio Sci.*, 38(3):1058, 2003.
- [35] Ao, C. O., "Effect of Ducting on Radio Occultation Measurements: An Assessment Based on High-Resolution Radiosonde Soundings," *Radio Sci.*, 42:RS2008, 2007.
- [36] Xie, F., et al, "An Approach for Retrieving Marine Boundary Layer Refractivity from GPS Occultation Data in the Presence of Super-Refraction," *J. Atmospheric and Oceanic Tech.*, 23:1629–1644, 2006.
- [37] Meehan, T. K., et al, "A Demonstration of L2C Tracking from Space for Atmospheric Occultation," *Proc. 2008 GNSS Meeting of the Institute of Navigation*, Savannah, GA, Sept. 16–19, 2008.

Remote Sensing Using Bistatic GNSS Reflections

Scott Gleason, Stephen Lowe, and Valery Zavorotny

16.1 Introduction

Remote sensing using GNSS-reflected signals is in many ways a logical extension of traditional radar remote sensing. However, there are a number of important differences that need to be considered before the environmental measurements obtainable using GNSS signals are deemed useful. The most important difference between GNSS and traditional remote sensing applications is that GNSS signals must be received and processed using a bistatic configuration. The transmitter and the receiver are not colocated, resulting in the transmitted GNSS signal reflecting or scattering in a “forward” direction, as opposed to the more conventional back-scattered signals used in traditional remote sensing applications. An illustration of the bistatic configuration, with the receiver on a LEO satellite is shown in Figure 16.1. Secondary differences from traditional approaches have to do with the GNSS signal characteristics suitability to surface sensing and the relatively low signal power levels transmitted by the GNSS satellites.

It is well known that GNSS signals scattered from ocean, land and ice are affected by the reflecting surface, and the changes induced by the surface can be observed. Understanding what exactly is being sensed and to what accuracy will drive the future applications of this novel technique. However, the existing research has shown that GNSS remote sensing has the potential to give environmental scientists a low-cost, wide-coverage measurement network that will greatly increase our knowledge of the Earth’s environmental processes. For this to be realized, the achieved accuracy of the measurements must be challengingly high. However, modest but still useful applications such as dangerous sea sensing and coarse surface characterization have already been demonstrated and have relatively easier requirements. In this introduction to the topic, we hope to illustrate both the opportunities and challenges of this technique, while inspiring others to become involved.

The reader is encouraged to use the references cited in this chapter and to consult some of the useful publications listed at the end of the chapter in the “Further Reading” section for more in-depth descriptions of the theory and results presented.

ubiquitous GNSS transmissions are often considered as part of the environment. Thus GNSS remote sensing is often called a passive, bistatic technique.

16.1.2 Remote Sensing Using Reflected GNSS Signals

Using Earth-reflected GNSS signals as a means of sensing the ocean surface was proposed as far back as 1988 by Hall and Cordey [4]. Some time later the concept was put forward as an alternative technique for ocean altimetry by Martin-Neira at the European Space Agency [5]. In 1998, the same principle was demonstrated as a useful tool to sense ocean surface roughness by Garrison and others [6]. Significantly, the first space-based detection of an ocean reflected GPS signal was achieved by Lowe et al, at NASA's Jet Propulsion Laboratory [7] using fortuitously acquired calibration data from the SIR-C radar experiment on board the U.S. Space Shuttle.

In addition to the advances mentioned above, there has been significant progress in other areas of GNSS bistatic remote sensing during the past decade. This includes the recovery of wind speed and direction using multiple reflected signals captured using an aircraft based instrument by Armatys in 2001 [8] and Garrison et al, in 2002 [9]. Alternatively, sea ice sensing was shown to be possible by Komjathy et al in 2000 [10]. Another area of pressing need is a better knowledge of the near surface soil moisture content for agriculture and urban planning applications. In this regard, the GNSS bistatic technique has shown to be very promising based on aircraft and platform results by Masters at the University of Colorado [11] and Katzberg et al, at NASA Langley Research Center [12], both in 2005. The ability to make altimetry measurements of the reflecting surface has been demonstrated repeatedly during various aircraft campaigns by researchers in the U.S. and Europe, as the experiments performed by Lowe et al in 2002 [13] and Ruffini et al in 2004 [14] have shown. Additionally, the feasibility of this technique for global ocean, land, and ice sensing at spacecraft altitudes was demonstrated by Gleason et al., in 2007 and 2008 using a relatively simple experiment configuration [15–17].

There have also been significant developments in the theoretical basis for this emerging technology, including the development of an advanced model for explaining the observed behavior of ocean scattered GPS signals. A widely used model based on the Kirchhoff approximation and geometric optics limit (KA-GO) [3] was put forward by Zavorotny and Voronovich in 2000 [18] and is often used in conjunction with the ocean wave spectrum developed by Elfouhaily et al, from 1997 [19] as a means of understanding the physical mechanisms behind the observed signal scattering. Additional models have also been proposed that have delved more deeply into specific areas, such as that of Thompson et al [20], which contains new insights into the predicted frequency and polarization characteristics of the reflected signals.

All of this work taken together makes a promising case for the future of this technology, which despite being nearly three decades old is still in need of much further study. Many of the results cited above are described in more detail within this chapter, including a hands-on example on the processing of GNSS reflections received from space.

16.2 Reflection Geometry

As a continuous wave of a GNSS signal meets the Earth’s surface, it scatters pseudorandomly off individual surface facets, resulting in a complicated interaction with the surface itself and significant mixing between individually reflected signal components. The detailed theoretical attempts to model this process are summarized in Section 16.4. Fortunately, the basic time and frequency characteristics across the surface are easy to predict (as will be explained below, predicting the directional power level after scattering is the more difficult aspect of the problem). Irrespective of the signal interaction with the surface, the signal time delay and Doppler frequency shift can be mapped across the surface accurately. An example of two individual reflection paths, both transmitted and received at the same point, is shown below in Figure 16.2.

Considering Figure 16.2, path 1 is shown passing through the point of specular reflection where the angles between the incident and reflected paths and the surface normal are equal, (θ) and path 2 is shown incident on the edge of the first isorange ellipse. The first isorange ellipse defines the surface ellipse where the delay of path 2 is exactly one chip (for example $\sim 300\text{m}$ for GPS L1 C/A code) longer than the distance traveled by Path 1. This ellipse is shown and labeled A in Figure 16.2. In addition to the distance traveled being longer at the edge of the isorange ellipse, the frequency will also shift due to the different viewing angles.

At spacecraft altitudes the size of the first isorange ellipse can extend up to and exceed tens of square kilometers. As one could expect, the range of time delays and frequency shifts of these signals will vary considerably. Determining how a GNSS signal spreads over the surface in the time and frequency domain is determined by relatively simple well-known equations, included in more detail in Appendix 16A on the DVD.

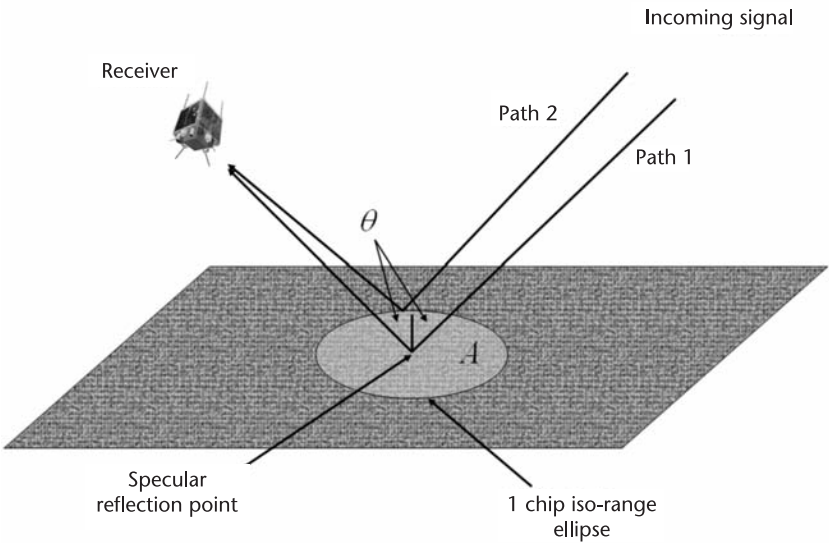


Figure 16.2 Surface representation of the bistatic scattering geometry showing two possible reflection paths.

16.2.1 Estimating the Surface Reflection Point Location

The point on the Earth that minimizes the transmitter-Earth-receiver distance is called the specular point and is used as the primary point of reference for measurements and modeling applications. A precise determination of this point is used to make an altimetry measurement and involves accurate modeling of the delay waveforms rising edge. Alternatively, it is used as the reference center for determining the estimated Doppler frequency shift and approximate code phase offset during signal search and acquisition. Surrounding the specular point the signal spreading in delay and Doppler can be mapped, and provides the basic reference frame around which ocean sensing models can be applied. Computing the specular point is described in more detail in an Appendix 16A on the DVD, with examples included on the DVD accompanying this coded in MATLAB/Octave.

16.2.2 Delay and Doppler Spreading over the Surface

The delay at every point on the surface can be calculated by simply tracing the path from the transmitter to the specular point to the receiver. Those surface points having a delay of one chip relative to the specular reflection define the extent of first isorange ellipse, shown as the area *A* in Figure 16.2. Additional lines map surface points having increasing delay in units of chips can be drawn over the complete region of received signal power. Similarly, the reflection from each surface point enters the receiver with a specific Doppler frequency, where lines of constant frequency form hyperbolas on the surface. The geometry of these so called iso-Doppler hyperbolas depend on the positions and velocities of the transmitter and receiver and the signal transmit frequency. This mapping of the path delays and Doppler frequencies over the surface is illustrated for two geometries in Figures 16.3 and 16.4 for a receiver in LEO.

For these examples, the receiver was assumed to be at an altitude of 680 km and the GPS transmitter at an altitude of 20,000 km with their orbital velocities directly towards each other. For the case of a 10 degrees incidence reflection (Figure 16.3) the tighter bunching of both the isorange and iso-Doppler lines is clearly evident when compared with a higher incidence reflection of 40 degrees (Figure 16.4). The effects of the changing geometry, and how they change the time and frequency characteristics of the detected signals, needs to be included in the theoretical modeling of the surface scattering, as described below. The total size of the glistening zone, defined as the surface region where reflecting facets are likely to scatter the signal towards the receiver, will depend on the roughness of the surface and increase as a function of incidence angle. However, the area useful for remote sensing consists of a significantly smaller region around the secular reflection point.

16.3 Signal Processing

In processing a GPS signal within a typical receiver, the incoming signal must be correlated with a locally generated replica chipping code with the appropriate offset and Doppler frequency shift as explained in the early chapters of this book. In the case of scattered signals, the delay and frequency response differ greatly from a

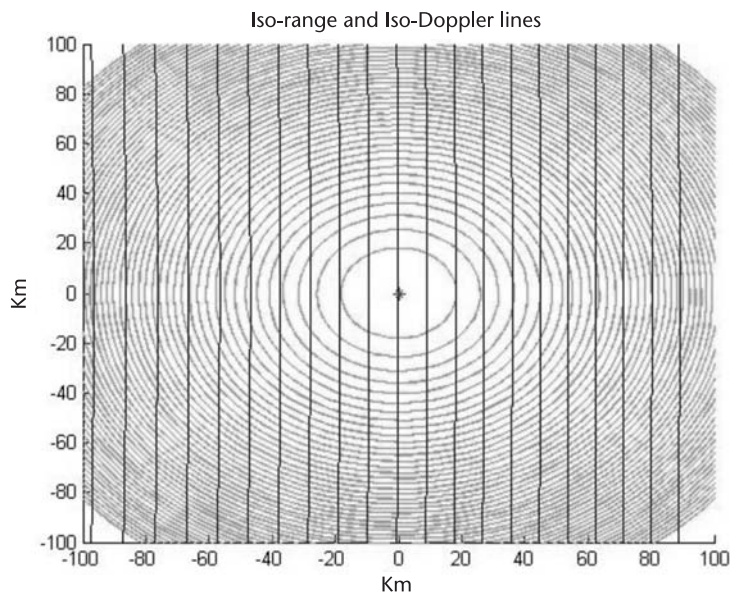


Figure 16.3 Surface isorange ellipses and iso-Doppler curves for the case of a 10 degrees incident reflection received from a LEO instrument. $\theta = 10$ degrees.

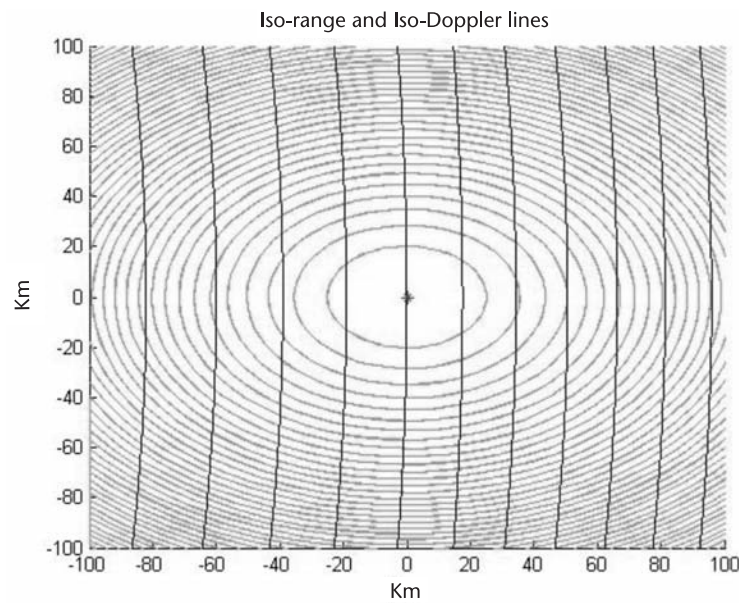


Figure 16.4 Surface isorange ellipses and iso-Doppler curves for the case of a 40 degrees incident reflection received from a LEO instrument. $\theta = 40$ degrees.

directly tracked signal, but the fundamental processing step is the same (and in some ways simpler).

The most basic inputs for a bistatic remote sensing system include data containing reflected signals and position, velocity and time information for the transmitting satellite and receiver. These inputs allow the recovery of reflection waveforms associated with an area on the Earth's surface. Using Figure 16.5 as a guide, the data processing steps can be followed towards the ultimate goal of remotely sensing a characteristic of the reflecting surface.

After the data has been processed, the result will be a representation of the detected signal power. This takes the form of a delay waveform, a delay-Doppler map, or a measurement of the bistatic radar cross-section. These basic measurements can then be compared with models or in situ data in an effort to connect the signal to a useful surface observable.

16.3.1 Detection and Surface Mapping

GPS signal processing begins by cross-correlating the data with a signal replica and coherently integrating, normally over 1 ms of data (determined by the length of the GPS PRN repeat sequence). By coherent its meant that the signal is processed

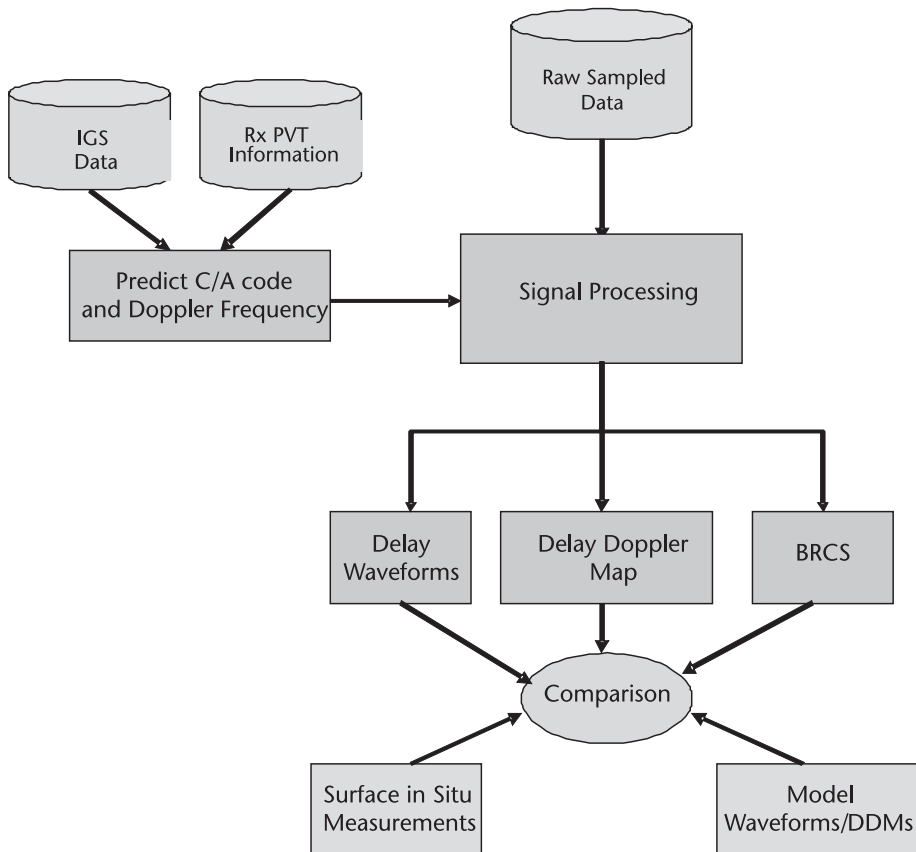


Figure 16.5 Data flow diagram of the basic processing functions.

using both its in-phase and quadrature signal components, with the possibility of computing a carrier phase angle based on these two values. However, all phase information is lost when the signal power magnitude is computed using the square of the in-phase I and quadrature Q signal components. Typically, the phase of the received signal is unpredictable and not related to the transmitted GPS carrier phase, making the signal magnitude the primary observable. Exceptions to this are when the surface roughness is small compared to the 19 cm L-band wavelength, or when the viewing angle is near grazing incidence where surface irregularities are shadowed and don't contribute.

Before any detailed processing can be started the signal needs to be detected at the delay and frequency near the specular reflection point. The two widely used methods for the initial acquisition of the reflected signal are listed below.

- Ideally, estimate the center frequency and time delay using known geometries and clock information. See the MATLAB/Octave scripts on the DVD for examples.
- Methodically search over all possible time delays and a complete or partial range of possible frequencies offsets, possibly using FFTs to perform correlations quickly. This is more time-consuming but eliminates the need for external information.

After the signal is coarsely located in delay and frequency, a range of delays and frequencies over the glistening zone are then processed using these initial values as a reference point. This step will be generally familiar to most readers, with the slight deviation that during processing for remote sensing there is nearly always a need to loop over a range or delays and frequencies (as opposed to traditional processing where a signal peak is tracked at a given delay and frequency). A block diagram of the code delay and frequency processing loops is shown in Figure 16.6.

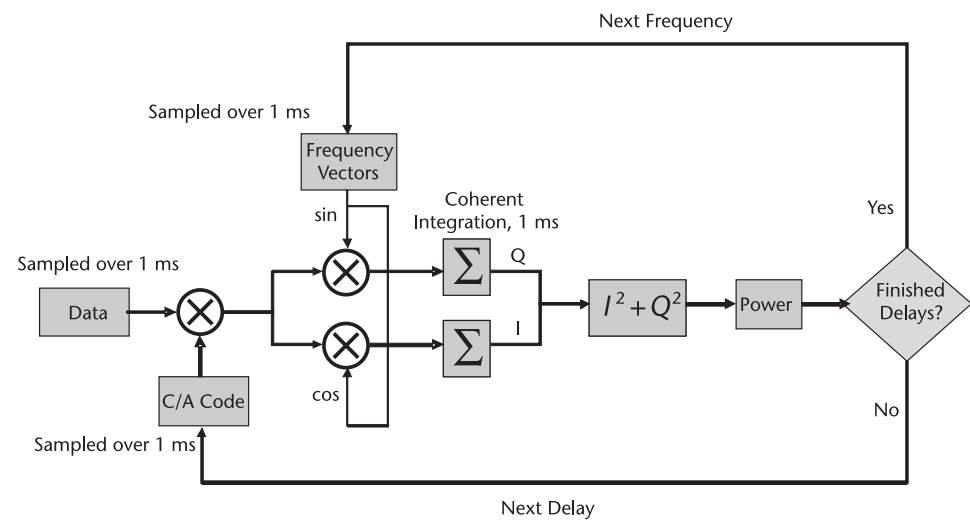


Figure 16.6 Flow diagram illustrating the basic processing loops for 1ms of sampled data.

Another important aspect of processing the signals for use in remote sensing is due to the surface scattering process. This interaction with a random surface makes it necessary to average consecutive correlations over time to uncover the true signal power profile. As the signal is averaged over time, while taking into account the changing dynamics, the overall scattered signal power can be mapped as a function of delay and frequency.

16.3.2 Averaging Consecutive Correlations

When a GPS signal is reflected from a rough surface the carrier phases of individual reflection points sum together in unpredictable ways at the receiver, resulting in a completely random received phase at the receiver. This causes the total received power level to fluctuate over time due to constructive and destructive interference between individual reflections. This phenomenon is known in traditional remote sensing applications as fading or speckle noise and is unavoidable for a diffusely scattering surface.

To reduce the effects of speckle noise, the signal must be summed or averaged over consecutive coherent correlations to extract a useable waveform in the presence of the speckle caused by the scattering surface. This necessitates that the whole process shown in Figure 16.6 be repeated over several consecutive milliseconds of raw data. For every trial delay and Doppler frequency, consecutive milliseconds are summed together as illustrated in Figure 16.7. This has the effect of mitigating the fading or speckle noise caused by the random scattering and results in a better estimate of the true signal power. (See [16] for a detailed analysis of signal fading statistics.) In traditional radar remote sensing applications, these separate summations are often referred to as “looks” at the surface, and are assumed to be statistically independent from one another.

Because the signals are summed over several looks, the effects of the system dynamics on the summation process need to be properly considered between

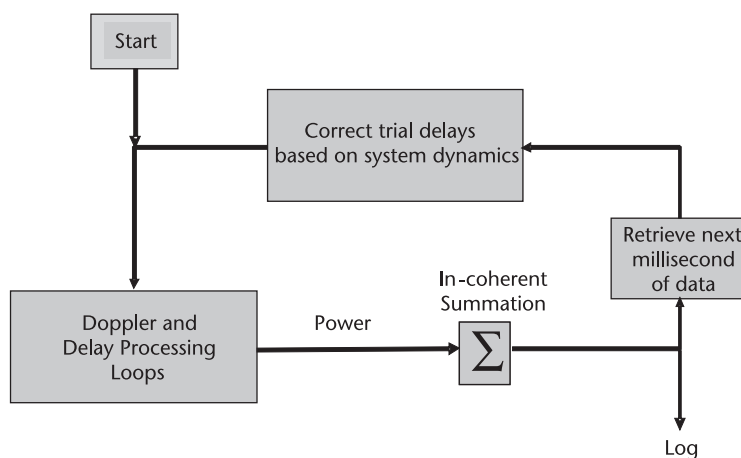


Figure 16.7 Block diagram of signal processing, including a series of noncoherent summation of coherent correlation power outputs. The process is done in parallel at each trial delay and Doppler frequency.

correlations or looks. From experience, it has been found that when the summation interval is limited to 1 second, the only term that has a noticeable effect on the shape of the returned signal is the first derivative of the code delay. However, additional terms could be included to increase precision.

Last, it is possible to integrate coherently over longer or shorter periods of time and then perform summations in a similar manner. However, the actual coherence time of the scattered signal will determine the optimum coherent processing interval. For signals received from space-based receivers a 1-ms coherent correlation was used (unless otherwise stated).

16.3.3 Delay Waveforms and Delay Doppler Maps

The effects of noncoherent summation described above can be demonstrated using any one of the ocean-reflected signals detected using the bistatic GPS experiment on the LEO UK-DMC satellite. A good example is that of the relatively strong signal collected from GPS satellite PRN 28 on March 23, 2004.

The delay waveforms (i.e., the center Doppler frequency is held constant during processing) for this signal and four different averaging intervals are shown in Figure 16.8. In this case the signal is only partially visible after a single look, or 1ms correlation. It can be seen that several additional looks are needed to observe the true underlying signal shape. The signal shape is distinguishable after both 10 ms

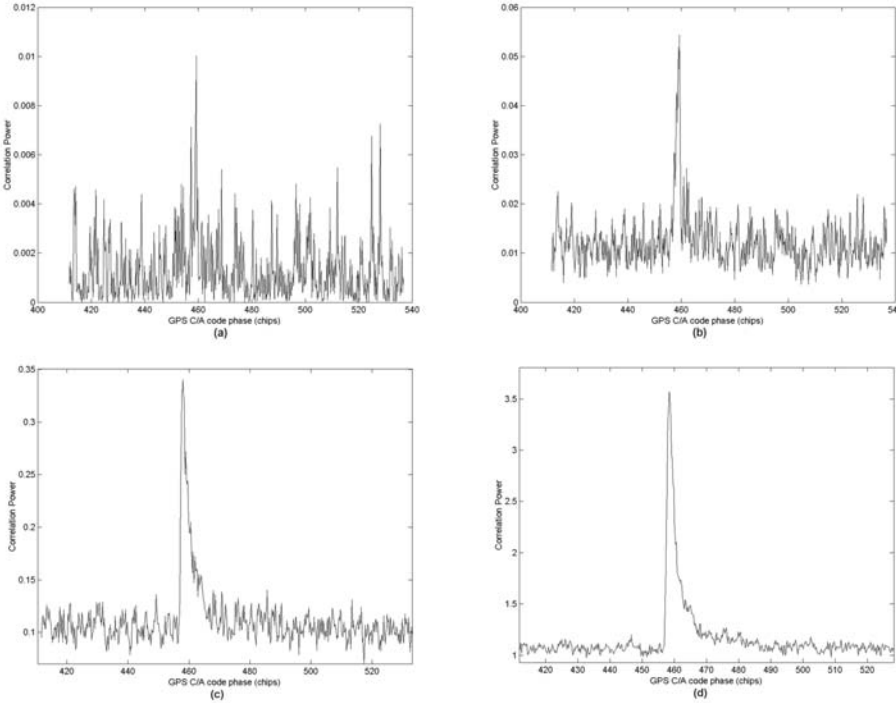


Figure 16.8 Signal found in data collected by the UK-DMC satellite on the 23 March, 2004, for GPS satellite PRN 28, using noncoherent integration times of: (a) 1 ms, (b) 10 ms, (c) 100 ms and (d) 1 second. (Source: [15]. Reprinted with permission, © IEEE.)

and 100 ms of summation and additional smoothing was achieved over a total duration of 1 second.

The signal power fluctuation over the averaging interval will vary considerably depending on the surface and is primarily due to speckle noise, but also includes other system noise sources, notably instrument thermal noise. Averaging of the signal also serves to increase the overall processed signal to noise ratio and makes the observed waveform easily identifiable above the noise floor. For the case of very weak signals (from very rough seas for example) this is an equally important aspect of the averaging process.

In order to produce a full delay-Doppler map of the scattered signal power it is necessary to process the data over a large range of delay and frequency bins. Figure 16.9 shows an example of a delay-Doppler map.

As the surface roughens, signal power will be returned from surfaces farther and farther from the specular reflection point. This causes the received power to spread over greater regions of time and frequency space. Notably, the spreading of the signal is not only influenced by the scattering surface but also by the time and frequency signal response functions. This second factor is what causes the resulting signal power map to slightly resemble a horseshoe, specifically, the 1-KHz frequency bandwidth associated with the 1-ms coherent processing interval. As the delay increases, corresponding to larger isorange ellipses, the largest intersection with regions bounded by iso-Doppler hyperbolas occurs at the two frequencies tangent to the ellipse. Thus, away from specular, most signal power comes from two Doppler frequencies surrounding the specular frequency, and as the delay

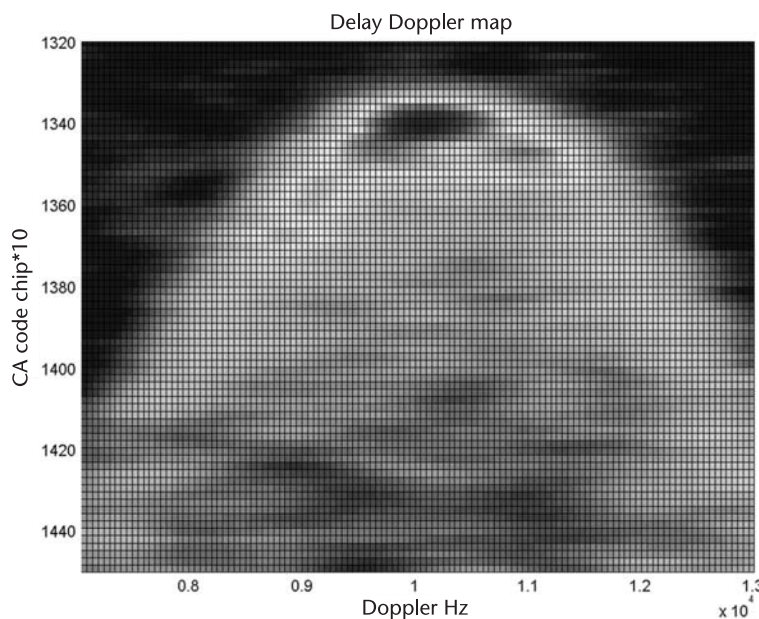


Figure 16.9 Delay-Doppler map of the ocean reflected signal of GPS satellite PRN 26, found in the May 21, 2004 data set collected by the UK-DMC satellite. The dark circular area at approximately (Hz = 10,000, CA Code = 1,340) is the area of highest detected signal power. (Source: [15]. Reprinted with permission, © IEEE.)

increases, so does the separation between the specular frequency and the two surrounding frequencies, thus the horseshoe shape (observe from Figures 16.4 and 16.3 that the a change of 1,000 Hz on the surface is well within the glistening zone).

An interesting challenge in using the delay-Doppler map is to extract directional surface information using only a single reflection. This technique has seen some preliminary success using aircraft data, as described in the work of Soulat [21].

16.4 Remote Sensing Theory

16.4.1 Bistatic Surface Scattering

In GNSS bistatic radar, the complex amplitude of the received signal (the voltage) is cross correlated with a replica of emitted signal over a coherent integration time, T_i . Frequently, this procedure is called a match-filter signal processing. The end result of this type of coherent signal processing is an ability to form a synthetic footprint that would ultimately determine the shape of the recorded waveform and the spatial resolution of the GNSS bistatic radar. For every epoch t_0 the code cross-correlation relative to the received signal u taken at a variety of delays, τ , can be expressed as the integral [18, 22]:

$$Y(t_0, \tau, f_c) = \int_0^{T_i} a(t_0 + t') u(t_0 + t' + \tau) \exp(2\pi i f_c t') dt' \quad (16.1)$$

Here T_i is the coherent integration time, and $a(t)$ is the replica of the PRN code sequence taking values of $\{-1, +1\}$ on a time duration τ_c . The coherent integration time T_i should be comparable or smaller than the coherence time τ_{cor} of the scattered field at the receiver point in order to perform the convolution procedure (16.1) with linear phase shift between replica $a(t)$ and signal $u(t)$. The oscillating factor containing f_c is meant to compensate for a possible Doppler shift of the signal $u(t)$ associated with this phenomenon. For signals received from spacecraft, the signal coherence time τ_{cor} , has been observed to be on the order of 1 ms [7, 16], while signals received from aircraft are known to remain coherent for considerably longer durations, on the order of 10–20 ms.

Only scattered waves with equal time delays and equal Doppler shifts could be successfully aligned with the code replica in order to produce maximum correlation according to (16.1), and it always happens within a so called glistening zone caused by a random distribution of the surface slopes. (See Section 16.2 for a quantitative analysis of delay and Doppler spreading over the surface and the glistening zone.) The size of the glistening zone is driven by the variance of surface slopes, where the larger the variance of surface slopes the larger the glistening zone extends across the surface.

The scattering toward the receiver is produced mostly by specular reflections from a statistical ensemble of large-scale (larger than several radio wavelengths) slopes of the surface. Therefore, the strongest scattered signal comes only from the

center of the glistening zone near the nominal specular point on the mean sea surface. Away from the glistening zone, the contribution from the quasi-specular reflections diminishes, eventually to be replaced with significantly weaker diffraction scattering from a small-scale surface component. Here we neglect this type of scattering as it is too weak to make a significant contribution to the total received signal power. Using this understanding for the physical scattering mechanism, we can apply a Kirchhoff theoretical model to estimate the expected scattering behavior [3]. In essence, we are combining a multitude of “smooth” reflection surfaces together to represent the signal scattering from a rough ocean surface, where every point on the surface is approximated with a local tangent plane.

The scattered GNSS signal $u(t)$ arriving at the receiver position \vec{R}_r can then be modeled by the integral taken over the mean sea surface [18]:

$$u(\vec{R}_r, t) = \int D(\vec{\rho}) a \left[\frac{t - (R_0(t) + R(t))}{c} \right] g(\vec{\rho}, t) d^2\rho \quad (16.2)$$

where $D(\vec{\rho})$ is the amplitude footprint of the receiver antenna; $a(t)$ is the GNSS signal PRN code; $R_0(t)$ and $R(t)$ are distances to the transmitter and the receiver, respectively, to some point $(\vec{\rho}, z = \zeta(\vec{\rho}, t))$ on the “smoothed” rough sea surface with an elevation of $\zeta(\vec{\rho}, t)$, fluctuating about the mean surface level. Over the individual local tangent planes the Earth’s curvature is neglected; $\vec{\rho} = (x, y)$; the transmitter and receiver positions are in the $x = 0$ plane, and z is a vertical axis or local surface normal.

The above analysis applies to the scattering of signals from surface components with spatial scales of several wavelengths greater than the incident carrier wavelength (i.e., the GPS L1 wavelength is ~ 19 cm). Alternatively, a contribution to scattering from surface components with spatial scales smaller than several radio wavelengths can be calculated separately using the perturbation theory. Additionally, serious limitations occur for scattering at low grazing angles and from very corrugated land and broken ice. In this case, more sophisticated scattering models that take into account multiple scattering and diffraction effects due to sharp edges is required [23, 24]. On the other hand, when surfaces are very even and flat, such as lakes and seas under low wind conditions, or first-year, young ice, the coherent component rises in the scattered GNSS signal.

In the Kirchhoff approximation, the function g describes propagation and scattering processes:

$$g(\vec{\rho}, t) = -V(\vec{\rho}) q^2 \exp \frac{ik(R_0(t) + R(t))}{4\pi i R_0 R q_z} \quad (16.3)$$

where V is the Fresnel reflection coefficient; $\vec{q} = k(\vec{n} - \vec{m})$ is the so-called scattering vector, where $k = 2\pi/\lambda$ is a radio wave number; \vec{m} is the unit vector of the incident wave; and \vec{n} is the unit vector of the scattered wave. Upon substituting (16.3) into (16.2), and then into (16.1), and assuming that integration over the accumulation time T_a is equivalent to averaging over a statistical ensemble of surface elevations

$$\langle |Y(\tau, f)|^2 \rangle = \frac{1}{T_a} \int_0^{T_a} |Y(t_0, \tau, f)|^2 dt_0 \quad (16.4)$$

After making some additional assumptions, we arrive at the bistatic radar equation [18]:

$$\langle |Y(\tau, f)|^2 \rangle = \frac{\lambda^2 T_i^2}{(4\pi)^3} P_t G_t \cdot A \cdot s_{1bit} \cdot \int \frac{G_r}{R_t^2 R_r^2} \chi^2(\tau, f) \sigma^0 dS \quad (16.5)$$

where:

T_i : the coherent integration time;

$P_t G_t$: Transmitter EIRP (effective isotropic radiated power);

A : Double-way atmosphere attenuation (−1.2 dB);

s_{1bit} accounts for any losses due to 1 bit sampling (−2 dB), for systems using 2 or more bit sampling this term can be ignored;

G_r : Receiver antenna gain;

R_r and R_t : distance between the specular point and the receiver/transmitter respectively;

χ^2 : Woodward ambiguity function (WAF), which describes the range and Doppler selectivity of the coherent radar;

σ_B^0 : normalized bistatic radar cross-section (BRCS) of the rough surface, which gives a portion of the scattered power carried by the outgoing plane wave in a specific direction, while the unit surface being illuminated by the unit wave incoming at another direction.

The WAF can be approximated by the square product of two functions: the triangularly shaped (in the case of GPS L1 signals) correlation function $\Lambda(\tau)$ and the sinc-shaped function $S(f)$. The first term determines an equirange annulus zone, and the second one determines an equi-Doppler-frequency zone. It is similar to the WAF used in the unfocused SAR technique [25]. The width of $\Lambda(\tau)$ is determined by a twice length of PRN code (or a twice chip length, $2\tau_c$), and the width of $S(f)$ is determined by the twice inverse of the coherent integration time, $f_{Dop} = 2/T_i$. For fixed positions of the transmitter and the receiver both WAF and BRCS are functions of reference surface S coordinates.

The characteristic shape of the two-dimensional waveform, or Delay-Doppler Map, emerges as a convolution of the WAF with the BRCS function σ_B^0 [21]. The WAF is close to unity within an area formed by the annulus zone and the Doppler zone, and tends to zero outside this area (note that for the new modulation schemes applied to new signals, such as BOC, this term needs to be re-evaluated). The geometry of these zones for two different elevation angles for a typical spacecraft receiver are shown in Figures 16.4 and 16.5. In applying this model to new GNSS signals, it would be necessary to substitute the appropriate delay and Doppler responses for the new signal characteristic. Chapters 2 and 17 discuss some of the new characteristics to be found in future GNSS signals.

16.4.2 The Bistatic Radar Cross Section

The effect of surface roughness is described by σ_B^0 , the normalized bistatic radar cross-section (BRCS). In the geometric-optics limit of the Kirchhoff approximation this term is represented by the following expression [3]:

$$\sigma_B^0 = \pi |V|^2 \left(\frac{q}{q_z} \right)^4 P \left(\frac{-q_{\perp}}{q_z} \right) \quad (16.6)$$

Though this value is a function of the scattering vector, \vec{q} , for fixed positions of the transmitter and the receiver above a surface, this vector can be regarded as a function of the coordinate $\vec{\rho}$ in the mean surface plane. The value of σ_B^0 depends on a complex Fresnel coefficient V which in turn depends on a signal polarization state, a complex dielectric constant of the reflecting medium, ϵ , and the local incidence angle.

Factor $P(\vec{s})$ in (16.6) is the probability density function (PDF) of large-scale “smoothed” surface slopes $\vec{s} = \nabla_{\perp} \zeta(\vec{\rho})$. Usually, the most probable orientation of surface slopes is parallel to the mean plane, $z = 0$. Then, the PDF has a maximum at $s = 0$, and the bistatic cross-section σ_B^0 has a maximum at $\vec{q}_{\perp} = 0$, i.e., at the nominal specular direction with respect to the mean surface. Note that the width of σ_B^0 in terms of ρ describes a glistening zone produced by quasi-specular points on the surface. MATLAB/Octave scripts are available on the DVD that demonstrate the application of the scattering model described in Section 16.4 as well as examples of generating ocean wave PDFs using real ocean observables, such as wind speed and direction as described below.

Some GNSS reflection receivers have the capability to sample the waveform only with respect to time delay, τ , while the frequency offset f is fixed and intended to compensate the Doppler shift associated with the nominal specular point on the Earth’s surface. In this case, we deal with a one-dimensional delay waveform, as shown in Figure 16.8. The leading edge of such a waveform up to the peak value is produced by the central elliptic annulus zone (filtered by the S function) when it expands from zero to its maximal value. The one-dimensional waveform forms a decreasing trailing edge after the peak because of the WAF behavior over time lags, and/or of the BRCS recession along radial directions according to the distribution of surface slopes. Because of the latter reason, the specific shape of the leading edge and an exact position of the correlation power peak is a function of surface roughness. For rougher surfaces, the leading edge is more stretched and the peak is more shifted toward later time lags. This means that accurate GNSS altimetry measurements depend on the availability of accurate information on surface roughness.

Equation (16.5) deals with values obtained by averaging over a limited number of independent samples. Such values themselves contain residual noise, which might affect our ability to accurately measure the average waveform. The issue of noise in waveforms and their impact on the accuracy of remote sensing of ocean wind and surface topography was addressed in Section 16.3.2. Equation (16.5) relies on the condition that $T_i < \tau_{cor}$. The coherence time can be estimated as

$\tau_{cor} = \rho_{coh} / v_r$, where ρ_{coh} is the coherence length of the scattered field at the reception point, and v_r is the velocity of the receiver. According to Van-Cittert-Zernike theorem, ρ_{coh} in the far zone increases linearly with the distance from the instantaneous footprint patch on a scattering surface. The size of the footprint patch, or in our case, an annulus zone, depends on the current time delay between the replica and the reflected signal. Therefore, a computation of the coherence time becomes a nontrivial problem which can be explored in more detail by consulting [26].

16.4.3 Sea Surface Modeling

The strength of the bistatically scattered signal from the ocean surface is mostly affected by the surface roughness since variations in salinity of the ocean is rather small. It is believed that for gradual and linear surface gravity waves the slope PDF $P(\vec{s})$ can be approximated by the anisotropic bivariate Gaussian distribution [18, 21, 27]:

$$P(\vec{s}) = \frac{1}{2\pi\sqrt{\det(M)}} \exp\left[-\frac{1}{2} \begin{pmatrix} s_x \\ s_y \end{pmatrix}^\dagger M^{-1} \begin{pmatrix} s_x \\ s_y \end{pmatrix}\right] \quad (16.7)$$

where matrix M is

$$M = \begin{pmatrix} \cos \varphi_0 & -\sin \varphi_0 \\ \sin \varphi_0 & \cos \varphi_0 \end{pmatrix} \cdot \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_c^2 \end{pmatrix} \cdot \begin{pmatrix} \cos \varphi_0 & \sin \varphi_0 \\ -\sin \varphi_0 & \cos \varphi_0 \end{pmatrix} \quad (16.8)$$

and φ_0 is the angle between the up-down wind direction and x axis, which is chosen here to lie within the incidence plane; σ_u^2 is an upwind mean-square slope (upwind mss); σ_c^2 is a cross-wind mean-square slope (cross-wind mss). $\sigma_{u,c}^2$ are wind-dependent and can be derived from a surface elevation spectrum $\Psi(\kappa)$ by integration over wave numbers κ smaller than a scale-dividing wave number κ_* . Sometimes, matrix M is called a directional mean-square slope in contrast to total mss which is defined as $2\sigma_u\sigma_c$.

One advantage of a Gaussian distribution is that the variance of slopes can be derived solely from a wave vector spectrum $\Psi(\vec{\kappa})$, of full surface elevations by integrating it over wave numbers, κ , which are smaller than a dividing parameter, κ_* :

$$\sigma_{u,c}^2 = \langle s_{u,c}^2 \rangle = \iint_{\kappa \leq \kappa_*} \kappa_{u,c}^2 \Psi(\vec{\kappa}) d^2 \kappa \quad (16.9)$$

One of the most popular models for the spectrum $\Psi(\vec{\kappa})$ is the model proposed by Elfouhaily et al. [19]. This model describes wind-driven waves in deep water under diverse wave age (often called *fetch*) conditions and agrees with the in situ observations of the first sun-glint derived wave slope measurements of Cox and

Munk [28], performed several decades ago. There are other situations when wind direction does not coincide with the maximum of the spectrum (e.g., when gravity waves undergo refraction on currents or on bathymetry, or waves generated by a local wind are superimposed with a swell). Such complicated scenarios are not considered here.

There exists a “three-lambda” heuristic criterion for the scale-dividing wave number κ_* proposed by Brown [29] based on fitting modeled curves for microwave back scattering cross-sections with cross-sections obtained in experiments with satellite radar altimeters. The same criterion was initially applied for use of the Kirchhoff approximation for the two-scale calculations of the bistatic cross-sections [18]. Later on, a reasonable $\kappa_* = 2\pi \cos \varphi / 3\lambda$ on the incidence angle φ was assumed in [30]. In the paper [31], an expression for κ_* is obtained that contains also a dependence on wind speed. It was obtained by fitting modeled curves for GNSS bistatic scattering cross-sections with cross-sections obtained in aircraft experiments.

Therefore, under favorable conditions the two-dimensional GNSS reflected waveforms allow us to map the glistening zone shape and orientation as determined by the PDF of L-band-filtered slopes of the ocean surface. The resolution of this map depends on several geometric and receiver parameters, such as the platform altitude and velocity, angle of the satellite elevation, the signal chip modulation length and coherent time integration of the receiver. Under well-developed oceanographic conditions (i.e., the waves and wind have reached an equilibrium), the statistical distribution of surface slopes would connect the wind speed and direction to the observed waveforms, as it follows from (16.8) and (16.9). The stronger the wind, the greater the variance of slopes and the wider the glistening zone becomes. The direction along which the corresponding mean-square slope (the size of the glistening zone) is maximal indicates the up/down wind direction.

Above we assumed Gaussian statistics of the ocean wave slopes. However, there are some indications that the actual PDF of slopes do not exactly follow a Gaussian shape at their tails [32]. In terms of the glistening zone, it implies that this departure affects a periphery of the zone. This would translate into some discrepancy for the value of the waveform, at relatively large time delays, τ , and large frequency offsets, f . The ability to discern the difference caused by the departure from the Gaussian PDF of slopes depends on residual noise of measurements for the tail area of the waveform.

From a single one-dimensional GNSS reflected waveform it is possible to retrieve the surface roughness and wind speed (under well developed conditions) [30, 33]. To illustrate the dependence of GNSS reflected waveforms on wind speed and wind direction let us consider a direct problem. According to the Elfouhaily et al model [19], an elevation spectrum of a well-developed wind-driven sea surface can be represented as a product of the radial, or omnidirectional, part of the spectrum, and the azimuthal part of the spectrum. The azimuthal part of the spectrum reproduces two main features of the directional spectrum: its anisotropy, or directionality, and the wave number dependence of the angular spectral width. The azimuthal part of the spectrum is a two-sided function; it does not distinguish between up- and down-wind directions. Therefore, within this model there is no

means to solve a directional ambiguity problem. However, real wave profiles have some degree of asymmetry with respect to the wind direction.

Calculations of one-dimensional waveforms for various geometries and wind conditions show that their sensitivity to wind is lower at low receiver altitudes than at high ones. This is because the size of the glistening zone becomes smaller and smaller for decreasing aircraft altitudes. In order to increase accuracy of the wind measurements from reflected waveforms it is amendable to keep aircraft altitude significantly high. However, as the altitude increases the waveforms become more and more stretched, requiring a larger range of time lags to see the wind dependence. This phenomenon has a negative outcome since larger time lags lead to coarser spatial resolution.

The simplest approach to use GNSS reflection waveforms for wave and wind retrievals would be to try to find the best fit between the model-generated waveforms for various wind speeds and directions and the measured ones. However, even though the solution of the direct problem shows a sensitivity of one-dimensional waveforms to the wind direction, the inverse problem of finding the wind direction is not easy to solve. Because of the symmetry of both σ_B^0 and $\Lambda(\tau)$ projected onto the Earth's surface, for the wind direction that forms angle φ_0 with the plane of incidence the resulting one-dimensional form is the same for angles $-\varphi_0$, $\pi - \varphi_0$, and $\pi + \varphi_0$ as well. Use of additional measurements at a significantly different azimuth angle could reduce this ambiguity to the level of up- to down-wind. One possible way to resolve this ambiguity is using calibration data from external in situ sources, such as ocean buoys. This technique and methods for resolving the ambiguity has been previously demonstrated for an aircraft experiment [8, 33].

16.4.4 Bistatic Scattering from Land

Bistatic scattering from land is characterized by lower average return and larger variability of the signal compared to GNSS signal scattering from the ocean surface. This takes place due to the large variations in dielectric constant ϵ and roughness for land. For the ocean, saline water and L-band radio waves, the dielectric constant corresponds to $\epsilon \sim 70 + i60$. The respective Fresnel power reflection coefficient is ~ 0.7 for normal incidence. Conversely, for average bare soil with average moisture content a typical dielectric constant is $\epsilon \sim 15 + i2$, with a Fresnel power reflection coefficient of ~ 0.3 , and even smaller for drier soils. The real part of the dielectric constant is responsible for the part of the electromagnetic energy that is reflected by the surface, and the imaginary part of the dielectric constant, which is related to the conductivity of the medium, is responsible for the part of the energy which is absorbed by the medium. Dielectric properties of land vary significantly depending on the geology present. For practical reasons, we limit our scope in this discussion to plain bare soils. Usually, it is a mixture of sand, clay, and silt [34], where the value of dielectric permittivity of the soil depends on water content.

Due to the relatively low reflectivity of soil a significant portion of the radiation can penetrate the surface and undergo further reflection and scattering from lower layers of soil. Understanding this layered dielectric property of soil, or dielectric profile, is important in the modeling of GNSS signal reflections from land.

Of course, the surface roughness remains an important issue as well. A rougher soil surface will produce a weaker bistatic signal within the main forward scattering lobe. This would make the task of soil-moisture retrieval difficult if the soil roughness factor is unknown. Even so, there have been attempts to describe PDFs for certain classes of land surfaces [35], but there is no universal PDF of land slopes because of the great variety of land compositions, geological and geophysical mechanisms and conditions that contribute to land surface topography over a large range of scales.

On the other hand, in contrast to the problem of ocean roughness, land roughness is not changing from day to day. Therefore, if the GNSS reflection technique is employed for regular monitoring of the same area of land, relative changes of the peak values of the correlation power with a proper calibration can be used for soil moisture measurements. This approach has been used in several GNSS reflection experiments aimed at soil moisture measurements [12, 36].

16.4.5 Bistatic Scattering from Sea Ice

Bistatic scattering from sea ice might differ significantly from ocean bistatic scattering mostly because sea ice roughness has a distinctly different character than that of waves on a wind-driven ocean surface [37]. The first distinction from an ocean surface is that for a frozen sea surface the Fresnel reflection coefficient is determined by the effective dielectric constant of ice, which varies in large margins and depends on various factors, such as origin, morphology, ice composition, age, thickness, presence of melt ponds on its surface, etc. [38]. For young thin ice the reflection coefficient might be affected by the dielectric constant of the underlying ocean water since the L-band signal penetrates through thin ice (to about ~30 cm). Additionally, the vertical structure of ice is rather complex and is formed by several processes [38]. The role of volume scattering at L-band in first year ice is limited by the fact that the spatial scales of the dielectric inhomogeneities are much smaller than the radiation wavelength [39]. Furthermore, volume scattering from a dry snow cover is expected to have a negligible effect at this wavelength.

The second important distinction of sea ice is its surface geometry. The first-year sea ice surface topography differs significantly from the multiyear, old ice topography. Roughness of the first-year sea ice is not a statistically homogeneous field, i.e., variances of height and slope may not be constant over the surface. Roughness of the first-year sea ice is mostly represented by broken ice accumulated at ice ridges which are not, in general, uniformly distributed over the surface. Presence of a significant portion of flat areas of unbroken ice makes the PDF of heights and slopes very different from a convenient Gaussian distribution. In [37] a model of the first-year ice surface PDF is proposed which contains a delta-function component attributed to flat areas having the same elevation and exponential tails with different decrements on each side of the delta-function. The conclusion was reached that such an anomalous behavior of the PDF is responsible for a rise of the coherent component in the altimeter reflected signal that cannot be described by a conventional incoherent scattering theory based on geometric optics limit of the Kirchhoff approximation. Even a stronger coherent effect of ice flat areas should be expected for lower-frequency forward-scattered GNSS signals. The strong

presence of a distinct coherent reflection component in data analyzed from space supports this theory [16].

16.5 Ocean Altimetry

16.5.1 Motivation

The primary scientific motivation for developing GNSS ocean altimetry is improving mesoscale eddy measurements. Mesoscale eddies represent one of the largest global climate modeling errors and they play an important role in the transport of momentum, heat, salt, nutrients, and other chemical properties of the ocean. They are ocean features analogous to atmospheric storms, which result in sea-height changes above the mean of about 10 cm on spatial scales between 10 and 100 km with temporal evolution of about one week to one month [40, 41]. Studies of eddies are essential to understanding the dynamics of ocean circulation on all space and time scales, including the large-scale meridional transport of heat by eddies and the mass transport by large-scale currents through eddy-mean flow interaction.

Global measurements of eddies have not been possible by orbiting radar altimeters such as Topex/Poseidon, Jason I, and OSTM/Jason II. This is due to the spatial and temporal scale of these measurements, which are on the order of 200 km and 10 days, respectively. GNSS ocean altimetry, on the other hand, has the potential to map the Earth in about one to two days and with 25-km spatial resolution. This is due to the large number of transmitters (GPS, Galileo and perhaps GLO-NASS). The passive instrumentation could also imply several inexpensive receivers on a single launch, such as the COSMIC atmospheric occultation mission where six GNSS receivers were placed on orbit using a single launch. The large number of transmitters and receivers allow multiple, simultaneous bistatic altimetry measurements compared to a single radar altimeter measurement. Figure 16.1 shows the geometry for a single receiver. The challenge for GNSS altimetry is to overcome the lower bandwidths and signal power compared to radar altimeters, which will likely involve using large receiver antennas.

Other possible applications of improved spatial and temporal global coverage include the monitoring of fast moving barotropic waves that propagate across ocean basins too quickly to be seen by the Jason 10-day repeat cycle [42, 43], and studies of shorter-scale variability in the coastal regions associated with wind-driven upwelling, flow perturbations by abrupt bathymetric features, and eddies and meanders generated by flow instabilities.

16.5.2 Aircraft Altimetry Measurements

The science rationale for GNSS ocean altimetry given in the previous section assumes global, space-based measurements. However, when research began on this topic in the late 1990s, there were essentially no GPS receivers on orbit that had the capability to even detect a GPS reflection. A search through archival SIR-C data did find two reflections [7]. These were used to create a model of the expected altimetric precision as a function of several experimental parameters such as receiver

height, reflection nadir angle, antenna gain, and GPS signal (L1 or L2), but an altimetric measurement was not possible with those data.

With essentially no opportunities to make GNSS altimetric measurements from space, several groups began performing aircraft measurements to assess the limiting precision and accuracy [13, 14, 44]. The same basic experimental setup was common to all these experiments (see [45] for one example). The aircraft was configured to have an up-looking antenna view the direct GPS signals and a down-looking antenna on the bottom of the aircraft to view the reflected signals. Modified GPS receivers sampled the antenna voltages at about 20 MHz, and these samples were recorded to large disk arrays. Additionally, the up-looking signal was processed by a science grade receiver to provide accurate aircraft positions. Post-processing involved processing the data with a software GPS receiver using the general principals described in Section 16.3 to detect the direct and reflected signals. For these experiments, the primary observable was the time delay between the direct and reflected signal arrival. An example of a direct and reflected signal retrieved from an aircraft is given in Figure 16.10. This delay, plus the receiver and transmitter locations are used to estimate the aircraft height over the ocean. With the aircraft location obtained from the onboard receiver, the height information then gives a profile of the ocean's surface topography along the specular path.

All experiments used the GPS L1 C/A code signal, which has a 1-MHz chipping rate, however Lowe et al. [13] processed the L1 and L2 P(Y) code signals by obtaining Y-code chips over the experiment data collection interval. The P(Y) code, with its higher 10-MHz chipping rate was shown to produce more accurate height measurements. The Galileo constellation is expected to have signals with even greater bandwidths, which may be the signals of choice for ocean altimetry.

The aircraft measurements and extrapolations to spacecraft altitudes identified several limiting errors for ocean altimetry which included white system noise,

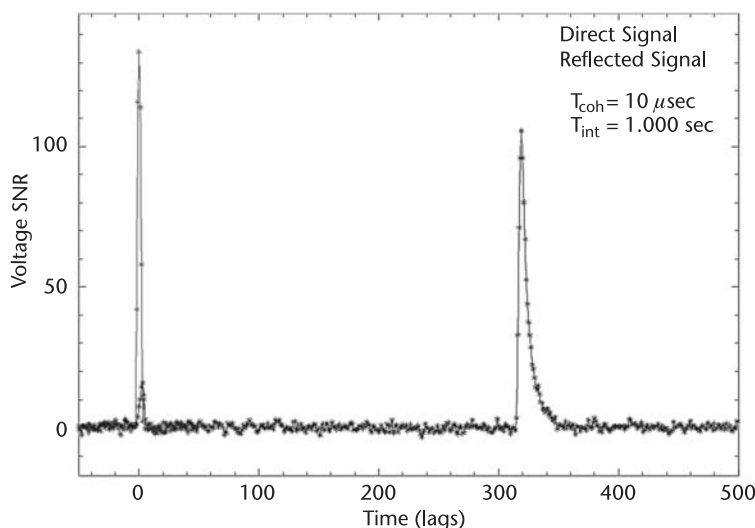


Figure 16.10 A typical direct and reflected GPS signal. The direct signal (on the left) is received on top of the aircraft, followed in time by the reflected signal (on the right) 200 lags (samples) later. The down-looking reflection antenna also detected the direct signal.

speckle, and mismodeling effects. The low-power GPS signals and the reflection's reradiation back up to spacecraft altitudes results in very weak signals there. A large-aperture antenna is needed to overcome white system noise; however, the antenna gain advantage is also limited by Rayleigh fading, or speckle from the surface. This implies the maximum signal-to-noise during a coherent processing interval is 1, and increasing the antenna size/gain beyond this limit has almost no effect on the final delay error (see [2, 16] for more details on the relationship between speckle noise and antenna gain).

The delay measurement between the direct and reflected signals requires determining the location of the specular point on the reflected signal's waveform. This requires an accurate model of the surface state, antenna gain pattern, hardware bandpass, and other parameters. In some aircraft data, the observed height for a given satellite is quite consistent, however, the height is systematically inconsistent between satellites over the same ocean surface, indicating that these models are not accurate enough over the large range of geometries seen in aircraft experiments. Although the situation may be easier in space, improving the models so that consistent and accurate heights are produced is an area for further research.

16.5.3 GNSS Ocean Altimetry from Space

Essentially all GNSS signals detected in space have used open-loop recorders to collect several seconds of data that are downloaded to Earth for processing [7, 15, 46]. Science-quality altimetry measurements will require orders-of-magnitude more reflections, which will require onboard processing of the reflection waveforms. Typical GPS receivers, which have three correlators in a DLL for each tracked satellite, do not have sufficient hardware for this task. These three correlators correspond to the top three points of the direct signal in Figure 16.10. A number of additional correlators, corresponding to the points on the reflected waveform in the figure, are required to obtain the reflected signal's waveform, especially the rising edge which is most important for altimetry. Models will be required to center about 10–100 correlators on the expected reflected signal's delay (note that waveforms from space are much wider and require more correlators than lower altitude aircraft measurements). For receivers implementing correlators in FPGA circuits, the additional hardware is essentially a replication of more correlator circuits on large FPGA chips.

It was noted that high-gain antennas will be required to detect the reflected signals in space, but the advantage of GNSS altimetry over traditional radar altimetry is the multiple, simultaneous reflections available. This implies an advanced antenna array is needed that can steer multiple high-gain antenna beams towards each specular reflection point simultaneously, as shown in Figure 16.1. The hardware driving such an antenna array will need the geometry of the transmitters, receiver, and the Earth, and for steering it will need to phase each antenna array element, for each formed beam. This implies the receiver itself should control the antenna array, as it already has all the necessary information. Further, the phasing circuit is essentially the same as a correlator circuit so it can be easily duplicated in a FPGA, and the phasing calculation requires similar geometric modeling to the delay calculation needed to window the reflection lags.

The two major additions to a typical GNSS receiver, multilag processing and multibeam antenna-array phasing, are being developed by several groups within ESA and NASA. It is expected that a flight receiver capable of processing reflections will be available in 2010. Because these advanced receivers are usually flown on science missions for orbit determination or atmospheric-occultation work (see Chapter 15), the addition of an antenna array somewhere on the spacecraft could allow GNSS reflections to be collected in parallel to the primary mission. This is similar to the paradigm used for the DMZ receiver to collect reflection data, but in this case the receiver will be capable of processing the waveforms onboard. The special difficulties associated with GNSS-reflection remote sensing, weak incoherent signals requiring special hardware, have so far precluded funding for a dedicated mission. Soon, there will be many inexpensive opportunities to collect a wealth of signals and assess the true potential of this concept.

16.6 Ocean Wind and Wave Sensing

16.6.1 Aircraft Wind and Wave Measurements

The first airborne experiments recording one-dimensional GNSS reflected waveforms performed wind speed measurements only. Wind-vector retrievals using one-dimensional reflected waveforms obtained simultaneously from three GPS satellites was demonstrated in [8, 33]. In Figure 16.11 the results of wind vector retrievals are presented in the form of arrows attached to the aircraft flight track. The evenly distributed arrows are QuickSCAT satellite scatterometer data.

It is obvious that for practical reasons the one-dimensional approach with multiple satellites is limited to relatively low-flying platforms when the size of annulus-generated footprints and distances between them are of the order of several km, or the measurements are taken close enough to be used together (i.e., not on completely different ocean regions as is the case for most space sensed signals). For an orbital (LEO) GNSS receiver this scale would translate into several hundred kilometers, over which most parallel reflections would be too far apart for use in deriving wave direction. For wind vector retrievals using orbital-based receiver, a more preferable approach would be utilizing a full two-dimensional waveform obtained from a single GNSS satellite as proposed in [21]. Notably, given the relatively long wavelength of the GPS carrier, wind sensing will be susceptible to wind/wave inconsistencies, which will occur often in dynamic wind conditions. For wind sensing to be reliable using GNSS signals, the ocean waves and wind must be at equilibrium or under conditions of well-developed seas.

In 2008, an approach was proposed by Cardellach and Rius [32] proposing a new algorithm for extracting the PDF of the sea surface slopes from GNSS reflected signals. According to this proposal, the PDF is discretized, and the unknowns to be solved are the set of discrete PDF values. An advantage of this new approach is that the resulting inversion system is linear, with no need for a priori or iterating sequences. The technique has been validated in three steps: the forward model has been checked under different geometries and sea surface conditions, generating waveforms with errors below 2%. Then a series of end-to-end simulations indicated that the technique could be applied to one-dimensional waveforms, or delay maps,

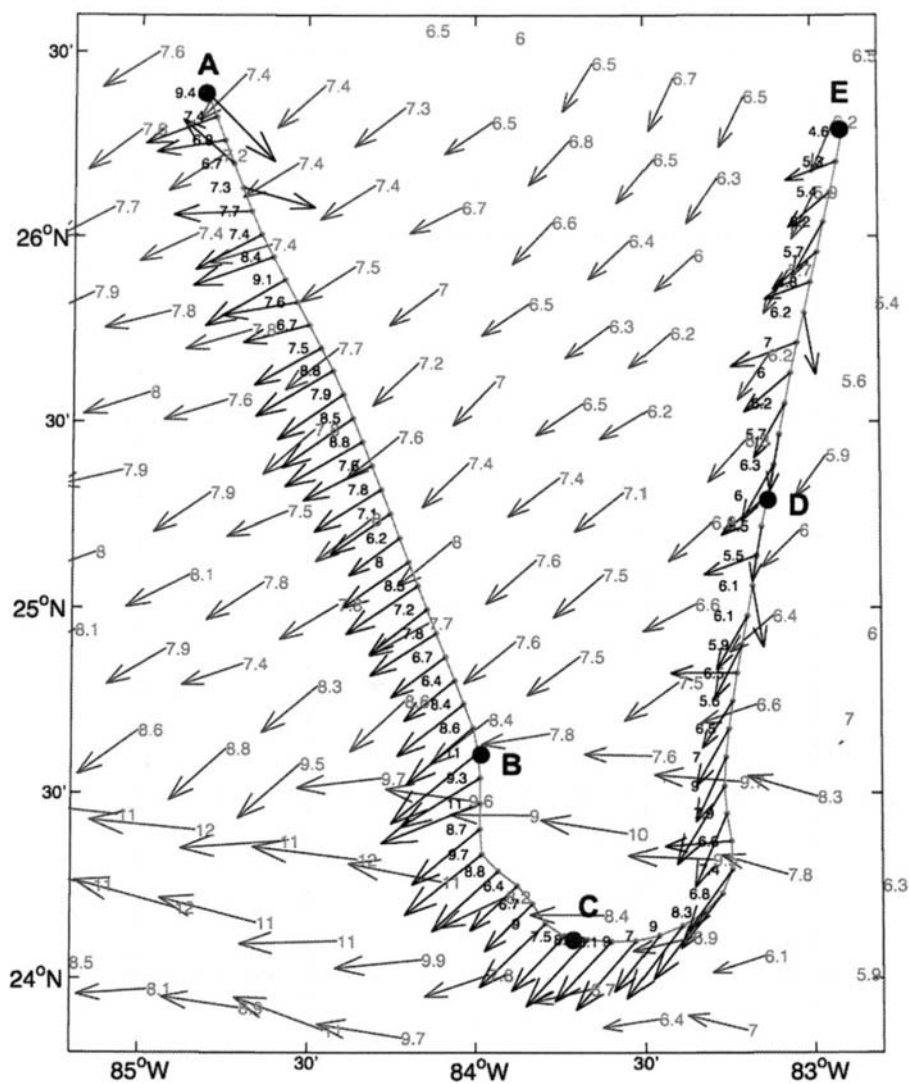


Figure 16.11 GPS-derived wind vector estimates at the end of the flight for Hurricane Keith on 1 Oct 2000 overlaid on QuikSCAT wind field measurements (Courtesy of Komjathy et al, [33]).

to obtain isotropic PDFs at 1% error. From two-dimensional waveforms, or delay-Doppler maps one can generate the so-called two-sided PDFs, obtained by combining of the PDFs retrieved separately from two sides of the glistering zone. The simulations show some deterioration of the retrievals as the elevation angle decreases or approaches to nadir. Finally, the algorithm has been tested on real data. It was found that, the retrieved PDFs reflect the effect of the wind direction, consistent with the ENVISAT SAR near-simultaneous observations. From here sensitivity to up-/down-wind wave asymmetry was found, which in its turn allowed breaking the 180° ambiguity characteristic for previous GNSS roughness, or wind, retrievals. The key results of this experiment are reproduced in Figure 16.12, with

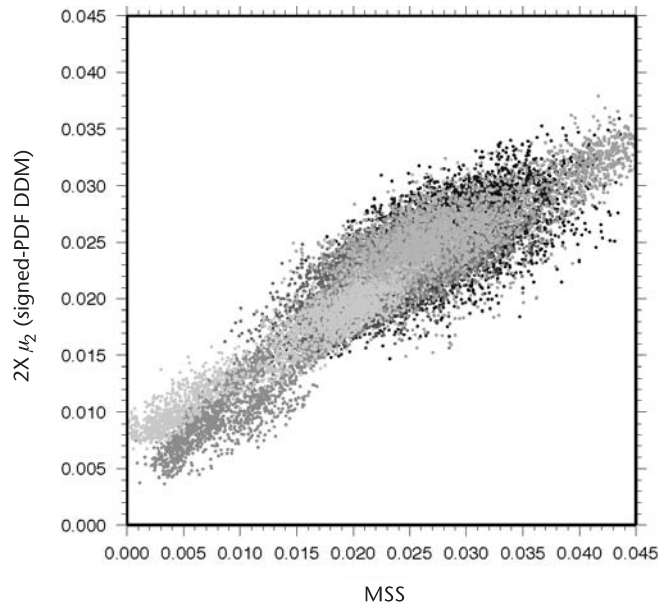


Figure 16.12 Sea surface mean square slopes computed using both Gaussian and non-Gaussian techniques for 1 second measurements made during 11 aircraft flights. Note that the new method demonstrated in [32] solves for the wave PDF itself (from which MSS can be derived) and has shown that the assumption of a Gaussian surface distribution does not always hold true (the two methods above correlated to 0.90). (Reprinted with permission of Estel Cardellach and Antonio Rius.)

permission and thanks due to Estel Cardellach, Antonio Ruis at IEEC in Spain and Remote Sensing of the Environment.

Additionally, the retrieval of sea surface directional mean square slope from GNSS reflection delay-Doppler-map (DDM) data was achieved during an experimental flight at 1-km altitude. This work involved processing the entire DDM to more precisely infer ocean roughness directional parameters. A thorough description of these results can be found in [21].

There have been numerous additional aircraft experiments performed over the past decades, the reader is encouraged to explore these publications independently [6, 8,9, 26, 30], including wind speed retrievals in Hurricane conditions [47].

16.6.2 Wave Sensing from Spacecraft

A demonstration that the signal power spreading in the frequency domain is connected to the sea surface has been carried out in [16] using several delay-Doppler maps retrieved from the spacecraft instrument aboard the UK-DMC satellite. In order to accurately estimate the Doppler spreading, the selected signals were averaged over a minimum of three seconds to obtain a clear determination of the power profile as a function of frequencies. The signals used in the experiment were collected over a representative range of surface mean square slopes and are listed in Table 16.1, together with colocated NDBC buoy information, the observed 3-dB Doppler spreading, incidence angle (theta) and other parameters.

Table 16.1 Doppler Spreading with Respect to NDBC Buoy Estimated Mean Square Surface Slopes

<i>Collection Date</i>	<i>GS PRN</i>	<i>Buoy MSS</i>	<i>Buoy Wind Speed ms</i>	<i>Buoy Wind Angle deg</i>	<i>Predicted 3dB Doppler (Hz)</i>	<i>Measured 3dB Doppler (Hz)</i>
October 31, 2005	20	0.0001	2.00	−67	1441	1300
March 4, 2005	27	0.0043	7.22	144	5430	3700
November 16, 2004	22	0.0075	8.76	−134	7216	4400
May 17, 2005	26	0.0094	10.83	130	8250	6100
September 3, 2004	17	0.0155	10.51	−42	9498	7500

In [27] Elfouhaily et al proposed a simple method of measuring the 3-dB Doppler bandwidth and relating it to the sea roughness. It was proposed that the 3-dB Doppler spectrum of an ocean reflected signal could be represented as,

$$BW_{3\text{ dB}} = 4\sqrt{\ln 2} \left(\frac{\sin \gamma}{\lambda} \right) \sigma_s |V^s| \sqrt{1 + \frac{\epsilon}{|V^s|^2}} \quad (16.10)$$

(See [27] for a detailed explanation of terms.)

As the UK-DMC delay Doppler maps are processed in terms of power (magnitude squared) a factor of $\sqrt{2}$ was used to convert the above equation to the equivalent 3-dB power bandwidth. This model was derived using a localized flat Earth and does not account for the spreading effects of the Earth’s curvature, which will be significant over the 3-dB glistening zone in most cases. Notwithstanding, Figure 16.13 shows the measured and predicted Doppler spreads for the five test signals and demonstrates a clear link between the wave MSS and Doppler spreading.

In this example the increasing Doppler frequency spread as the sea roughens is quite clear. These points are a good indication of the promising potential of this method over a wide surface area containing relatively uniform ocean wave conditions. The increasing upward trend is evident in both the measured and predicted signal widths, but with a noticeable gap between the measured and predicted values. This disagreement needs to be investigated further, but are believed to come mainly from Earth curvature effects and the variation of the receiver antenna pattern across the surface.

16.7 GNSS Bistatic Land and Ice Sensing

This section is based largely on an article [17] that appeared in the October 2008 issue of *GPS World* [17] and in the proceedings of the Institute of Navigation’s GNSS conference in Fort Worth Texas in 2006, reused with permission and thanks to *GPS World* and the Institute of Navigation [48]. Included on the DVD accompanying this book are the space-based data set and processing tools used in this chapter.

16.7.1 The History and Applications of GNSS Land Reflections

The earliest research into using GPS reflections for sensing land surfaces has been undertaken primarily at the University of Colorado [11] and NASA Langley

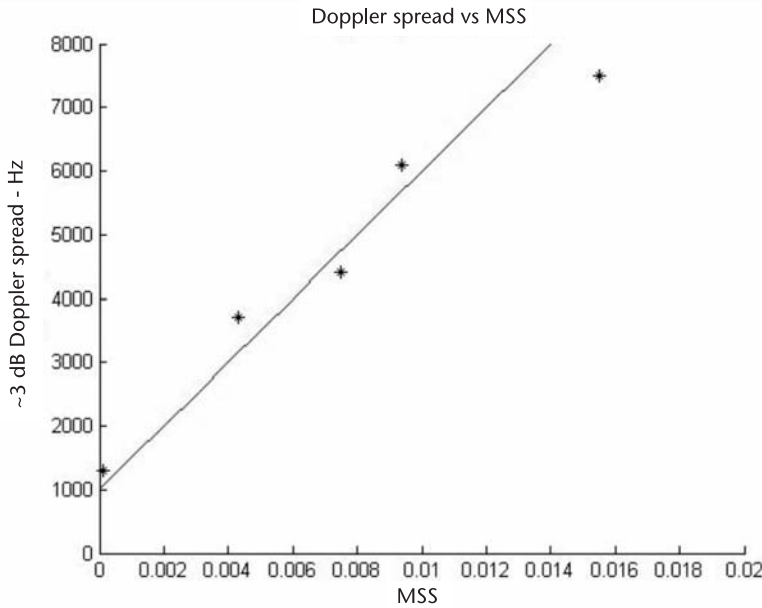


Figure 16.13 Doppler frequency spread versus buoy estimated wave mean square slopes. The measured 3-dB bandwidths are shown in black and the predicted in gray. A linear fit was done for each set.

Research Center [12]. These experiments were performed using measurements taken during aircraft flights and from Earth fixed platforms and have demonstrated both the possibilities and difficulties of sensing land characteristics with this technique. A very useful summary of remote sensing using land reflected GPS signals is the Ph.D. dissertation by Dallas Masters from the University of Colorado at Boulder [11]. The results presented in his research showed that GPS surface reflections were responding to surface moisture and displayed sensitivity to field boundaries and differences in land cover.

This could prove a significant discovery as soil moisture sensing is emerging as an important and often lacking parameter in numerous fields of scientific research. Some of the known applications of soil moisture measurements include; better determination of crop yields, flood prediction, drainage and run-off estimation (useful for urban planning), inputs into weather prediction models and in the study of chemical and nutrient transport in environmental studies. Additionally, the L-band frequency of GNSS signals has been shown to be sensitive to surface soil moisture, which provides an additional motivation. However, sensing soil moisture and/or terrain cover from LEO is expected to be challenging, as explained in Section 16.4.4, specifically due to the larger scattering region (or glistening zone) and mixing of terrain roughness and surface cover.

In the following sections, a demonstration of Earth reflected signals received from a space-based instrument are presented which clearly demonstrate the global potential of this technique attainable using satellite instruments. A more thorough examination of additional land reflections received from a LEO instrument can be found in [16].

16.7.2 Spacecraft-Detected Land Reflections

On December 7, 2005 a raw data set 20 seconds in length was collected using the bistatic GPS experiment on the UK-DMC satellite over North America, near Omaha City Nebraska. This data collection contained two satellite reflection points within the antenna surface footprint, both of which were open loop tracked across the entire data collection at the locations shown in Figure 16.14. Interestingly, it is known to have snowed in the days before the data collection.

The data was processed across several C/A code delays and Doppler frequencies as described in Section 16.3 above, using the Open Source software receiver included on the DVD accompanying this book. To mitigate the effects of speckle noise, an inherent effect of diffuse surface scattering, the single 1ms correlations or “looks” were summed together into 100 ms combined measurements across the entire duration of the data collection. The variations in signal magnitude are plotted in Figures 16.15 and 16.16 as a function of time for the two reflections.

The geometries between the two reflections will have a small impact on the overall signal magnitude, but over this short interval general comparisons are reasonable. Over the collections interval the fluctuations in signal magnitude are believed to come mainly from the scattering surface.

From examining the fluctuations in the reflected signal with respect to the terrain several interesting links can be observed. The first is the obvious spike in the signal magnitude as the reflection point crosses the Missouri river (second 2 in Figure 16.15). As the signal passes over water it would be expected to increase in magnitude, hence the suspicion that the increase in received signal power at second 12 in Figure 16.15 may be due to the presence of rivers to the West of Omaha. In Figure 16.16, there is a slight increase in signal power at second 10 as

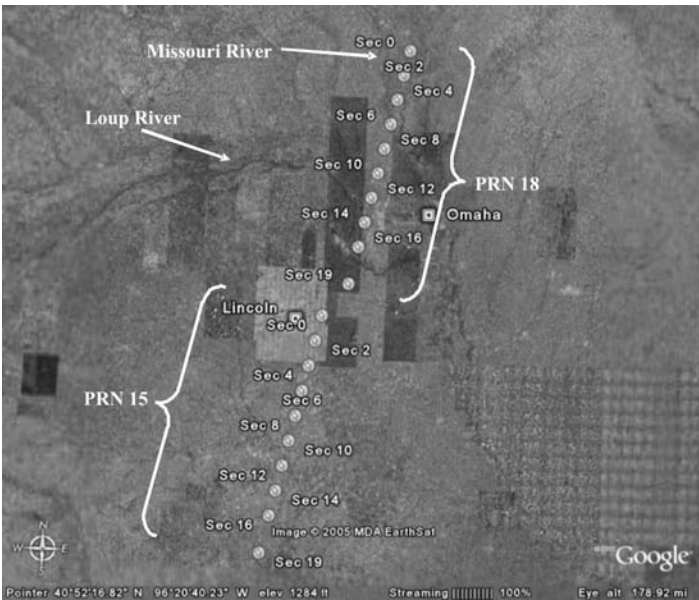


Figure 16.14 Google Earth image of the December 7, 2005 reflection locations. For GPS satellite PRN 15 (to the South) and PRN 18 (to the North).

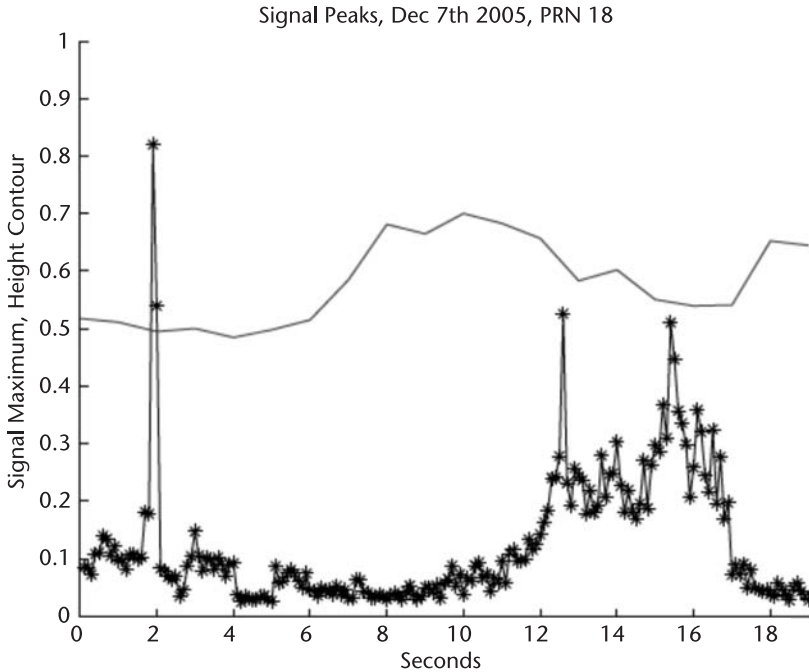


Figure 16.15 Signal maximums and height contour over 19 seconds of data, December 7, 2005, GPS satellite PRN 18 (Eastern Nebraska). A general normalized height contour (as indicated by Google Earth) is shown as a reference. Additional land collections from the UK-DMC showed weak correlation with the surface contours. The spike at second 2 is the signal crossing of the Missouri river.

the reflection passes to the east of Beatrice, Nebraska, this could also be due to the presence of a nearby river, but the exact cause is unknown and better in-situ ground information would be needed to confirm this.

16.7.2.1 Summary

Following this very preliminary demonstration a more thorough validation campaign needs to be carried out involving data collections over large areas in parallel with in situ measurements. These measurements could then be studied with respect to reliable in situ ground truth information and compared to models to determine if the reflections are responding to a useful surface observable, such as soil moisture or surface vegetation cover. For examples of successful retrievals of land characteristics the reader is encouraged to consult [11, 12].

16.7.3 The History and Applications of GNSS Ice Reflections

An accurate knowledge of the coverage and thickness of the Earth's sea ice is a key input parameter into determining the extent of global warming. Measurements from GNSS reflections could be used to help monitor the seasonal variations in ice around the polar regions, where good global sea ice information is critical to

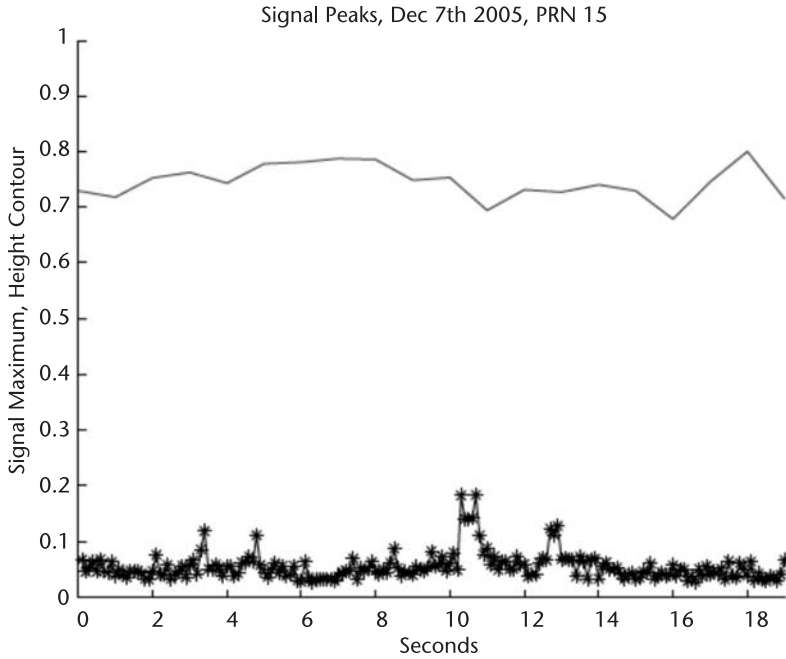


Figure 16.16 Signal maximums and height contour over 19 seconds of data, December 7, 2005, GPS satellite PRN 15 (Eastern Nebraska). The slight increase in signal power after second 10 corresponds to the reflection passing near Beatrice, Nebraska.

understanding the changing climate. In addition to climate change there are other applications and possible end users concerned with the Earth’s ice coverage. These include the sea ice impact on shipping routes, which is of crucial interest to sea navigators and the global shipping industries [49].

It has been shown that scatterometers, such as the SeaWinds instrument, can be used to remotely sense sea ice, including tracking icebergs and monitoring the ice shelves of Greenland [50]. This would lead one to believe that using GNSS signals in a bistatic scatterometry configuration to sense the Earth’s cryosphere is also possible. Following the initial work performed in [10], Maria Belmonte-Rivas at the University of Colorado at Boulder demonstrated that ice reflected GNSS signals contain information useful in the determination of ice classes, by measuring the ice permittivity and roughness [51].

The relatively low cost and wide coverage of the GNSS bistatic technique would make it a good candidate to contribute alongside the more advanced established instruments (such as the Canadian Space Agencies RADARSAT satellite) in observing the Earth’s cryosphere. Promisingly, a limited number of data sets collected using a satellite based instrument have shown that this technique is also applicable from LEO satellites, as described below.

16.7.4 Spacecraft-Detected Sea Ice Reflections

As of 2007, nongrazing reflected signals have been detected off sea ice on two occasions using a space-based instrument. On the 4th of February 2005, 7 seconds

The second collection was captured again by the UK-DMC instrument, consisting of 9 seconds of data near Antarctica in the Southern Ocean. This collection targeted an ice shelf off the coast and is shown in Figure 16.18.

The conditions at the specular reflection point locations at the time of the data collections were determined using data from the U.S. National Ice Center (<http://www.natice.noaa.gov/>). The National Ice Center data indicates that as the February 4th (Alaska) reflection moved across the ice surface it remained within a region of 9/10ths total concentration of first year ice, between 30- and 70-cm-thick for the entire 7 seconds. Where the total concentration is a measure of the amount of sea surface covered by ice as a percentage. A concentration of 9/10ths for floating ice (or very close ice in this case) is slightly below that of 10/10ths compact or consolidated ice sheets. In simple language this means that the region in question was over 90 percent covered with very densely packed ice flows, with possible small gaps between them. In contrast, the NIC ice data for the June 23rd collection (Antarctica) revealed that the sea was covered with between 8/10ths and 7/10ths ice. This presence of less compact sea ice resulted in a noticeably different reflected signal.

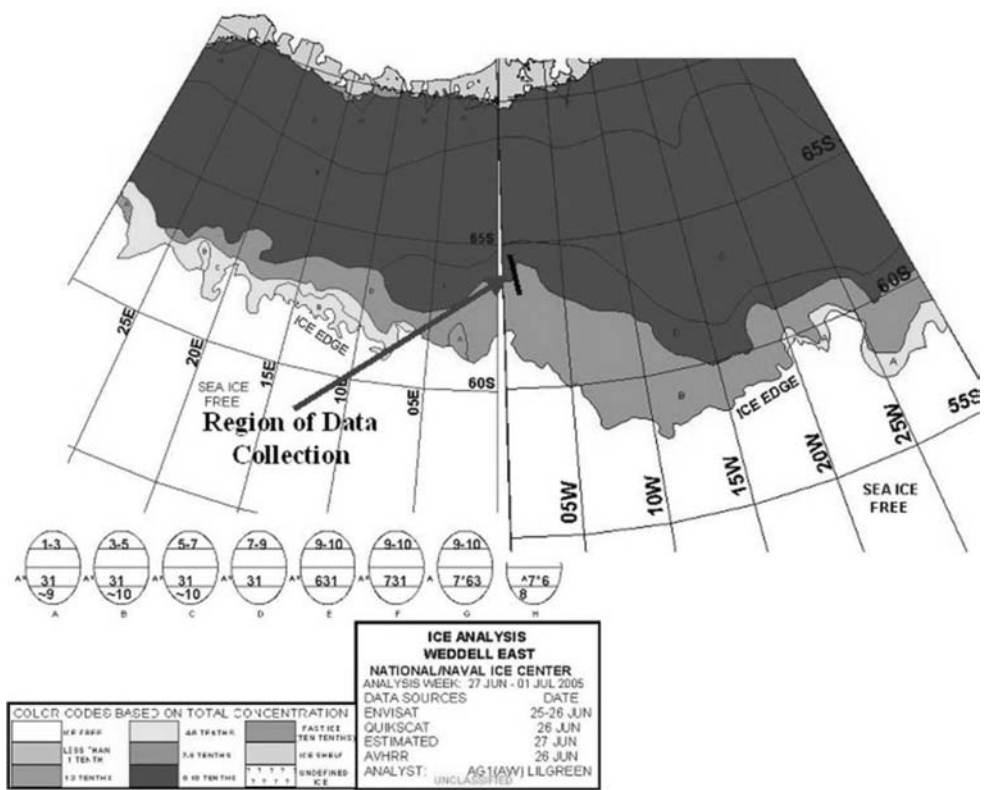


Figure 16.18 U.S. National Ice Center data indicating that the Antarctic ice shelf extended over the reflection location during the week after June 23, 2005. (North ~ down). Location of the GPS reflection is shown as a short black line. (NIC image courtesy Pablo Clemente-Colon and John Woods.)

16.7.4.1 Signal Response from Different Total Ice Concentrations

A comparison can be done using the February 4th data from 9-10ths sea ice and the June 23rd data where only 7-8ths sea ice was present to demonstrate that there is a noticeable difference in the signal magnitudes and spreading in delay of the postprocessed signals.

Interestingly, the February 4th signal reflected from a higher ice concentration had a consistently stronger peak (even with less averaging) and showed less spreading in delay and Doppler. Examples of the two signals power waveforms as a function of delay are shown in Figures 16.19 and 16.20 where the lower signal power levels and greater spreading in delay is clearly evident for the June 23 signal of lower ice concentration. To better observe the power spreading in the June 23 signal, one second of averaging was performed.

The greater spreading in delay observed in the June 23, 2005, signal is most likely due to surfaces reflecting power towards the receiver from distances away from the point of specular reflection. In other words the greater presence of water surfaces could be resulting in a larger total glistening/scattering zone. However, the roughness of the reflecting sea ice will also have an effect on the correlation magnitude and spreading (in addition to the increased presence of open water).

Additionally, the unique geometry of each reflection will alter slightly the magnitude and the spreading observed. The February 4th signals were observed at roughly 30 degrees incidence and the July 23rd signals at 19 degrees. However, it is believed that the difference in signal magnitudes between the two signals is due to primarily to the reflecting surface and only slightly distorted by their different incidence angles.

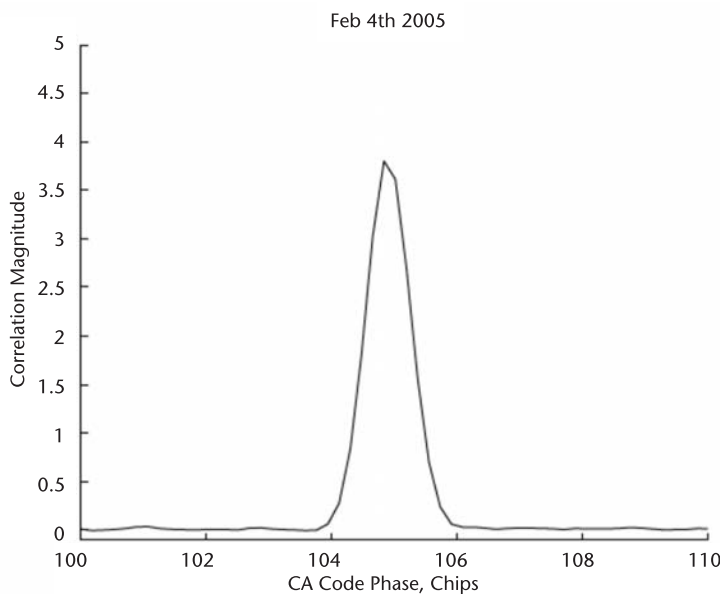


Figure 16.19 February 4, 2005, sea ice reflected signal delay waveform.

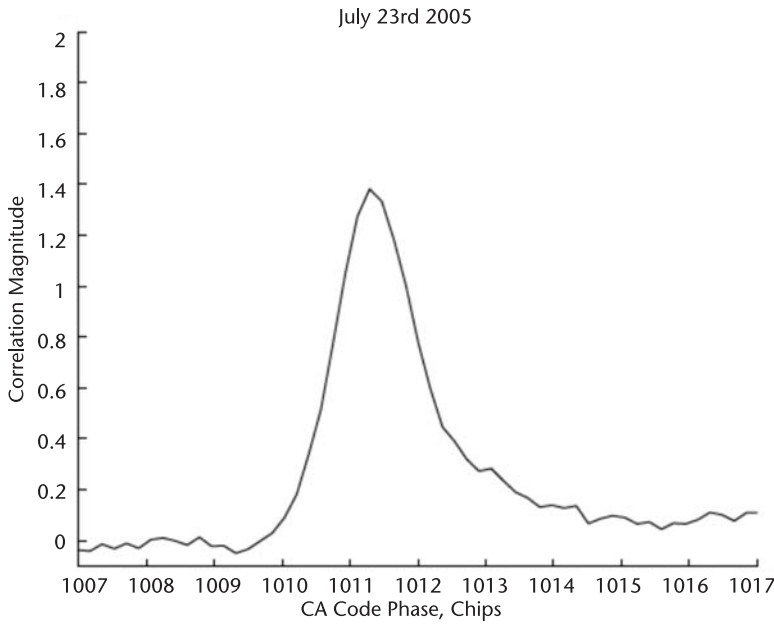


Figure 16.20 The June 23, 2005, sea ice reflected signal delay waveform.

16.7.4.2 Summary

The results shown above demonstrate that GNSS signals can be detected in LEO after reflecting from sea ice. Additionally, the different power returns and spreading in delay from different total ice concentration reveal this as a promising technology in sensing the sea ice concentration. However, as in the case of land sensing, a more thorough validation campaign is needed to accurately determine what is being sensed and what useful information may be contained in the signal power returns.

16.8 Data Provided on the DVD

A great deal of data and processing tools has been included on the book’s DVD, which will allow readers to repeat and even potentially improve on some of the results presented in this chapter. A summary of what has been included is described below and can be found on the DVD in subdirectory, `//Chapter16_Bistatic/`.

All of the data and code provided as part of this chapter is intended for reuse and distribution and released under open source licenses, please see license files in individual directories on the DVD.

16.8.1 Specular Point Calculation Scripts

The calculation of the specular reflection point of a GNSS reflection using a MATLAB/Octave script can be found in `//Chapter16_Bistatic/SpecularPoint Calculation`. Receiver and GPS satellite location information has been taken from

the UK-DMC Ocean based reflection data described below. An example calculation can be performed by running the `specular_point_demo.m` script.

16.8.2 Surface Scattering Model

An implementation of the Zavorotny/Voronovich bistatic surface scattering model can be found in `//Chapter16_Bistatic/ScatteringAndWaveModels/ZVModel/`. The implementation of this model was written by Maria Paola Clarizia and Scott Gleason. This model uses receiver and satellite locations and bidirectional surface wave slopes as inputs and generates a complete simulated delay-Doppler map over a specified glistering zone.

The sea surface wave model of Elfouhaily et al, capable of estimating bidirectional surface wave slopes with respect to surface wind conditions has also been included in `//Chapter16_Bistatic/ScatteringAndWaveModels/ ElfWaveSpectrum/`. The MATLAB implementation of this wave model was written by Dallas Masters.

16.8.3 Spacecraft Data and Processing Tools

The following space-based data collected from the UK-DMC satellite and processing tools have been included in `//Chapter16_Bistatic/SoftwareReceiver_Data/`. The contents of the individual subdirectories are listed as follows.

- A C software receiver (Daaxa) designed to process UK-DMC raw data collections;
- Example processing scripts used to configure the software receiver;
- MATLAB scripts used to plot the outputs of the software receiver;
- A raw UK-DMC data collection over the ocean, November 16, 2004;
- A raw UK-DMC data collection over land, December 7, 2005;
- A raw UK-DMC data collection over sea ice, February 4, 2005;
- GPS telemetry file with receiver PVT information for each collection, an IGS file for each GPS satellite for each collection and a navigation timing file are included for each of the three raw data sets.

Please consult the `ReadMe_SoftwareReceiver_Data_Discription.txt` document on the DVD for detailed notes on processing the data provided.

Special thanks are due to all those at Surrey Satellite Technology Limited involved in the design, implementation and operation of the UK-DMC bistatic experiment. In particular, we would like to acknowledge Martin Unwin, Mounir Adjrad, and James Wilhelm for their significant contributions. Additionally, we would like to thank Dr. Jeff Ward and Sir Martin Sweeting for granting permission to release this data publicly.

16.9 Further Reading

- For details on GNSS bistatic remote sensing, space-based results, please see [7, 16].

- For additional information on GNSS bistatic land sensing, please refer to [11, 12].
- References [10, 17, 51] provide further information on GNSS bistatic ice sensing.
- See [8, 9, 21] to read further on GNSS bistatic remote sensing, ocean wind, and wave sensing.
- References [2, 3, 13, 14] offer additional information on GNSS bistatic remote sensing, altimetry applications.
- References [2, 3, 18, 19] offer more thorough explanations of remote sensing theory and additional details on the theory discussed in Section 16.4.

References

- [1] Robinson, I. S., *Satellite Oceanography: An Introduction for Oceanographers and Remote Sensing Scientists*, Praxis Publishing Limited, 1994. Reprinted in 1995 and 1997 by John Wiley and Sons in association with Praxis Publishing Limited.
- [2] Ulaby, F. T., R. K. Moore, and A. K. Fung, *Microwave Remote Sensing: Active and Passive; Volume II: Radar Remote Sensing and Surface Scattering and Emission Theory*, Norwood, MA: Artech House 1982.
- [3] Bass, F. G., and I. M. Fuks, "Wave Scattering from Statistically Rough Surface," *International Series in Natural Philosophy*, v. 93, C. B. Vesecky and J. F. Vesecky (eds.), Pergamon Press, Oxford, U.K., 1979.
- [4] Hall, C., and R. Cordy, "Multistatic Scatterometry," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Edinburgh, Scotland, 1988.
- [5] Martin-Neira, M., "A Passive Reflectometry and Interferometry System (PARIS): Application to Ocean Altimetry," *ESA Journal*, Vol. 17, 1993. U.S. Patent 5 546 087, Aug. 13, 1996.
- [6] Garrison, J. L., S. J. Katzberg, and M. I. Hill, "Effect of Sea Roughness on Bistatically Scattered Range Coded Signals from the Global Positioning System," *Geophysical Research Letters*, Vol. 25, No. 13, pp. 2257–2260, July 1, 1998.
- [7] Lowe, S., et al, "First Spaceborne Observation of an Earth-Reflected GPS Signal," *Radio Science*, Vol. 37, No. 1, 2002.
- [8] Armatys, M., *Estimation of Sea Surface Winds Using Reflected GPS Signals*, Ph.D. thesis, University of Colorado, 2001.
- [9] Garrison, J. L., et al, "Wind Speed Measurements Using Forward Scattered GPS Signals," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 40, No. 1, Jan. 2002.
- [10] Komjathy, A., et al, (2000), "Sea Ice Remote Sensing Using Surface Reflected GPS Signals," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 24–28 July, 2000, pp. 2855–2857.
- [11] Masters, D., *Surface Remote Sensing Applications of GNSS Bistatic Radar: Soil Moisture and Aircraft Altimetry*, Ph.D. thesis, University of Colorado, 2004.
- [12] Katzberg, S. J., et al, "Utilizing Calibrated GPS Reflected Signals to Estimate Soil Reflectivity and Dielectric Constant: Results From SMEX 2," *Remote Sensing of Environment* 100, 17–28, 2005.
- [13] Lowe, S., et al, "5-ms-Precision Aircraft Ocean Altimetry Using GPS Signals," *Geophysical Research Letters*, Vol. 29, No. 10, 2002.
- [14] Ruffini, G., et al, "The Eddy Experiment: Accurate GNSS-R Ocean Altimetry from Low Altitude Aircraft," *Geophysical Research Letters*, Vol. 31, No. 21, June 2004.

- [15] Gleason, S.T., et al, "Detection and Processing of Bistatically Reflected GPS Signals from Low Earth Orbit for the Purpose of Ocean Remote Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 43, No. 6, June 2005, pp. 1229–1241.
- [16] Gleason, S., *Remote Sensing of Ocean, Ice and Land Surfaces Using Bistatically Scattered GNSS Signals From Low Earth Orbit*, Ph.D. thesis, University of Surrey, United Kingdom, 2006.
- [17] Gleason, S., "Reflecting on GPS: Sensing Land and Ice From Low Earth Orbit," *GPS World*, October 2008.
- [18] Zavorotny, V. U., and A.G. Voronovich, "Scattering of GPS Signals from the Ocean with Wind Remote Sensing Application," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 38, No. 1, March 2000, pp. 951–964.
- [19] Elfouhaily, T. M., et al, "A Unified Directional Spectrum for Long and Short Wind-Driven Waves," *Journal of Geophysical Research*, Vol. 102, No. C7, July 15, 1997, pp. 15781–15796.
- [20] Thompson, D. R., T. M. Elfouhaily, and J. L. Garrison, "An Improved Geometrical Optics Model for Bistatic GPS Scattering from the Ocean Surface," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 43, Issue 12, December 2005, pp. 2810–2821.
- [21] Soulat, F., *Sea Surface Remote Sensing With GNSS and Sunlight Reflections*. Ph.D. thesis, Universitat Politècnica de Catalunya, Spain, 2004.
- [22] Parkinson B. W., et al (eds.) *Global Positioning System: Theory and Applications*, Vol. I & II, AIAA, Washington, D.C., 1996.
- [23] Elfouhaily, T. M., and C.A. Guérin, "A Critical Survey of Approximate Scattering Wave Theories From Random Rough Surfaces," *Waves Random Media*, Vol. 14, pp. R1–R40. 2004.
- [24] Warnick, K. F., and W. C. Chew, "Numerical Simulation Methods For Rough Surface Scattering," *Waves Random Media*, Vol. 11, pp. R1–R30, 2001.
- [25] Elachi, C., *Spaceborne Radar Remote Sensing: Applications and Techniques*, New York: IEEE Press, 1988.
- [26] You, H., G., et al, "Stochastic Voltage Model and Experimental Measurement of Ocean-Scattered GPS Signal Statistics," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, No. 10, 2004, pp. 2160–2169.
- [27] Elfouhaily T., D. R. Thompson, and L. Lindstrom, "Delay-Doppler Analysis of Bistatically Reflected Signals from the Ocean Surface: Theory and Application," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 40, No. 3, March 2002.
- [28] Cox, C., and W. Munk, "Measurement of the Roughness of the Sea Surface From Photographs of the Sun's Glitter," *J. Opt. Soc. Am.*, Vol. 44, 1954, pp. 835–50.
- [29] Brown, G. S., "Backscattering From a Gaussian-Distributed, Perfectly Conducting Rough Surface," *IEEE Transactions on Antennas and Propagation*, Vol. AP-26, pp. 472–482, 1978.
- [30] Garrison, J., et al, "Wind Speed Measurements Using Forward Scattered GPS Signals," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 40, No. 1, Jan. 2002, pp. 50–65.
- [31] Thompson, D. R., T. M. Elfouhaily, and J. L. Garrison, "An Improved Geometrical Optics Model For Bistatic GPS Scattering From the Ocean Surface," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 43, No. 12, 2005, pp.2810–2821.
- [32] Cardellach, E., and A. Rius, "A New Technique to Sense Non-Gaussian Features of the Sea Surface From L-band Bi-Static GNSS Reflections," *Remote Sensing of Environment*, 112, 2008, pp. 2927–2937.
- [33] Komjathy, A., et al, "Retrieval of Ocean Surface Wind Speed and Wind Direction Using Reflected GPS Signals," *J. Atmos. Ocean. Tech.*, 21(3), 2004, pp. 515–26.
- [34] Njoku, E., and D. Entekhabi, "Passive Microwave Remote Sensing of Soil Moisture," *J. Hydrology*, Vol. 184, 1996, pp. 101–129.

- [35] Dierking, W., "RMS Slope of Exponentially Correlated Surface Roughness For Radar Applications," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 38, No. 3, May 2000, pp. 1451–1454.
- [36] Masters D., P. Axelrad, and S. Katzberg, "Initial Results of Land-Reflected GPS Bistatic Radar Soil Moisture Measurements in SMEX02," *Remote Sens. Environ.*, Vol. 92, 2004, pp. 507–520.
- [37] Brown, G. S., "A Theory For Near-Normal Incidence Microwave Scattering From First-Year Sea Ice," *Radio Science*, Vol. 17, No. 1, 1982, pp. 233–243.
- [38] Carsey, F. (ed.), *Microwave Remote Sensing of Sea Ice*, *Geophys. Monogr. Ser.*, Vol. 68, AGU, Washington, D.C., 1992.
- [39] Winebrenner, D. P., L. D. Farmer, and I. R. Joughin, "On the Response of Polarimetric Synthetic Aperture Radar Signatures at 24-cm Wavelength to Sea Ice Thickness in Arctic Leads," *Radio Science*, Vol. 30, No. 2, 1995, pp. 373–402.
- [40] Chao, Y., et al, "Modeling the Gulf Stream System: How Far From Reality?," *Geophys. Res. Lett.*, 23, 3155–3158, 1996.
- [41] Smith, R. D., et al, "Numerical Simulation of the North Atlantic Ocean at 1/10°," *J. Phys. Oceanogr.*, 30, 1532–1561, 2000.
- [42] Chao, Y., and L.-L. Fu, "A Comparison Between the TOPEX/Poseidon Data and a Global Ocean General Circulation Model During 1992–93," *J. Geophys. Res.*, 101, 14, 965–24, 976, 1995.
- [43] Stammer, D., C. Wunsch, and R.M. Ponte, "2000: De-Aliasing of Global High-frequency Barotropic Motions in Altimeter Observations," *Geophys. Res. Lett.*, 27, 1175–1178.
- [44] Rius A., et al, "Sea Surface State Measured Using GPS Reflected Signals," *GRL* 29 (23), Dec. 2002.
- [44] Lowe, S. T., et al, "A Delay/Doppler-Mapping Receiver System for GPS-Reflection Remote Sensing," *IEEE Trans. Geosci. and Remote Sensing*, 40(5), 1150–1163, 2002.
- [46] Meehan, T., et al, "Observations of GPS Sea Surface Reflections from the SAC-C Spacecraft," *Workshop on GNSS Oceanography with GNSS-Reflections*, July 2003, Barcelona, Spain.
- [47] Katzberg, S. J., O. Torres, and G. Ganoë, "Calibration of Reflected GPS for Tropical Storm Wind Speed Retrievals," *Geophysical Research Letters*, 33, L18602, doi:10.1029/2006GL026825, 2006.
- [48] Gleason, S., "Land and Ice Sensing From Low Earth Orbit Using GNSS Bistatic Radar," *Proceedings of the Institute of Navigation*, Fort Worth, Texas, September 2006.
- [49] Hamer, M., "The Polar Road to Riches," *The New Scientist*, January 21, 2006.
- [50] Long, D. G., "Land and Ice Applications of SeaWinds Data," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium 2000*, Vol. 3, 24–28 July 2000, pp. 1220–1222.
- [51] Belmonte-Rivas, M., *Bistatic Scattering of Global Positioning System Signals From Arctic Sea Ice*, Ph.D. thesis, University of Colorado, 2007.

New Navigation Signals and Future Systems in Evolution

Anthony R. Pratt

17.1 The History of GNSS

The following sections introduce a brief history of satellite navigation with reference to openly available satellite systems specifically designed and launched to support terrestrial navigation. In fact, almost all satellite navigation constellations have been designed and launched with a specific military objective. The Navy Navigation Satellite System (NNSS), popularly known as Transit, was designed to support the Trident missile system; GPS was designed as a military system to support land, sea, and air forces. During the development of these systems, the civilian applications became so beneficial that the United States provided access to sufficient signal characteristics that civilian users could receive and utilize the location and navigation capabilities provided by these systems. In contrast, Galileo was designed as a civil satellite navigation system but with an authorized or governmental signal/service as an additional capability.

It is hoped that the insight gained into the evolution of these systems will be seen by readers as providing a clear path to future capabilities with clear limitations provided by signal bandwidth, noise background, and the overcrowding of the ITU protected radio navigation satellite systems bands. The overcrowding is a result of the widespread recognition of the civilian and military benefits available from precise knowledge of location—a concept that few would have accepted only a quarter of a century ago. Such is progress!

17.1.1 GPS

The U.S. Global Positioning System (GPS) came into existence slowly during the late 1970s with just a few satellites operating in various test roles, as a replacement for the first generation NNSS (Transit) system. This phase continued until the 1990s when the constellation was built up for an initial operational capability (IOC) on December 8, 1993, and then a final operational constellation (FOC) was declared on April 27, 1995. The constellation itself went through several evolutions in concept until the current one was confirmed with 24 operational satellites and three active spares in six orbital planes, each with an inclination angle of 55° to the equatorial plane. Only two frequencies were used for the downlink from satellite

to user community designated L1 (1575.42 MHz) and L2 (1226.60 MHz) until 2009 when a third frequency L5 (1176.45 MHz) was added. The carrier frequencies, and all other signals, are derived from a common atomically stabilized clock operating at 10.23 MHz (as observed on the ground after compensation for relativity effects). The transmission at L1 is at the 154th harmonic while L2 is the 120×10.23 MHz and L5 is 115×10.23 MHz. The bands, where the signals appear, are protected by the International Telecommunications Union (ITU) for radio navigation. The signals in the L2 and L5 bands are shared with other services including aeronautical navigation radars, DME, TACAN and JTIDS. DME (distance measuring equipment) is widely used in civil aircraft navigation. The protected radio navigation bands and their occupancy by GPS, Galileo, and GLONASS satellite navigation systems are shown in Figure 17.1.

GPS, as the first of the modern satellite navigation systems, is primarily a military system providing two precision positioning signals, known as the P-code, on L1 and L2 and a Clear Acquisition signal (C/A code) on L1. The L2 signal can also be modulated with a C/A code component but this has rarely been activated. The P-code transmissions contain sections of a 38-week-long code with a chipping rate of 10.23M chips per second. Acquisition of the code signal requires a replica to be held in time alignment in the receiver circuitry. Without aiding, in the early stages of GPS receiver development, the search process to find the correct replica alignment was thought to be difficult to achieve. However, modern techniques have made this possible using multichannel receivers aided with high-accuracy clocks and innovations in the search procedures. The original application of the C/A code was to provide a simple means to access current GPS system time through the satellite data broadcast (it is contained in the second word in every navigation message subframe). Chapter 5 provides a description on how retrieve GPS time from the GPS L1 signal.

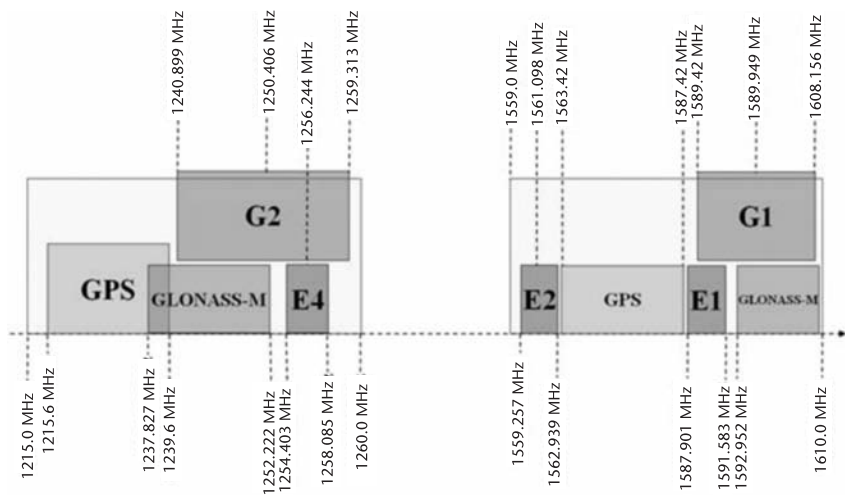


Figure 17.1 ITU frequency bands and application.

17.1.2 Modulation of Satellite Carrier Signals

Satellite navigation signals are modulated from a variety of different sources within the navigation payload. These are described as follows:

- Low-rate navigation data possibly encoded by an error correction system (such as a Viterbi forward error corrector) or added parity data bits. In data format, this is known as data modulation or as symbol modulation if encrypted with an error correction system;
- A (high-rate) spreading code, which causes the spectrum to be significantly widened, provides separate identity for each satellite signal on the same frequency and provides a processing gain for the signal during receiver processing;
- A spreading symbol, transmitted for every element of the spreading code sequence. Early GPS signals used simple spreading symbols known by the term *rectangular* to maintain the phase of the carrier component at a constant level for the duration of each spreading code element or chip.

There has been innovation in each of these groups and also in one further aspect of the transmitted signal—that is the way in which many different signals are combined on a single carrier. GPS signals (from satellites up to and including the Block IIA satellites) on the L1 carrier used carrier phase quadrature to provide for two different signals (C/A and P(Y) codes signals), separately accessible, to coexist on the same carrier. Since both the signals are binary, each signal is modulated by binary phase shift keying (BPSK) onto each of the two quadrature carrier components forming a quadrature phase shift keyed signal (QPSK). Later satellites, which include M-code signals, use more complicated schemes, such as Interplex, to provide for additional signal component transmission.

17.2 Motivation for Evolution

The first signs that additional satellite navigation signals might appear is generally attributed to the European Union (EU) in the 1990s when a number of options had been proposed. Several exploratory contracts were performed seeking to establish technical characteristics and market opportunities for an alternative to GPS. The original motivation, by popular understanding, arose from a desire to express the new sovereignty of the European Union. Paying for such a system required that specific budget sources could be found to support the development and deployment costs. In the European Union, a purely civil source of funding was established to support transportation (from the E.U. Transport Council). The anticipated costs of the system from inception have also evolved as the required complexity has become clear. Estimates in the range 5–10 B€ are now recognized as necessary.

The first technological activities in developing the signal structure began with an ad hoc committee known as the frequency task force comprised of experts provided by the member states (MS) of the EU at the expense of the MS, under

the chairmanship of Jean-Christophe Martin (F). This was quickly replaced by the signal task force (STF) under the same arrangements but with a broader focus upon the complete signal structure for the European Union's emerging system—Galileo.

An inner group comprising of experts drawn primarily from three MS (France, Germany and the United Kingdom) has been responsible for most of the overall signal design. The key signal characteristics were decided in small ad hoc meetings called as required, and the results subsequently proposed to the STF for debate and acceptance. Many options were considered and the few that have survived detailed scrutiny now constitute the Galileo Signal Structure.

During a similar time frame, experts in the United States were also exploring the potential for additional signals from the GPS constellation. First to gain acceptance was a new transmission in the lower L-band at 1176.45 MHz—now known as L5. This is a QPSK transmission with a 10 Mchip/sec chipping rate similar to that of P(Y) code signals in L1 and L2. The intention was to support aeronautical users. New code structures were introduced with both primary and secondary codes. The primary codes using selected sections from products of 8191 length m-sequences with repetition intervals of 1 ms (length 10230). The secondary codes are the same for all satellites and are selected Neuman-Hoffman sequences (one each for the in-phase and quadrature components).

Subsequently, further consideration was given to the provision of a continuous civil signal on L2—as the L2 C/A code was rarely transmitted. The primary motivation was to support high-accuracy land survey users who would benefit from the higher effective carrier-to-noise ratios associated with a receiver able to fully despread the L2 civil code signals. There are two code types now in use, known as CM and CL. The code lengths are 10230 and 767500 corresponding to repetition intervals of 20 ms and 1.5 seconds at a chipping rate of 511.5 kchips/sec. The codes are transmitted in time multiplex. Details of the new transmissions in GPS are freely available and can be found in the relevant Interface standards (IS GPS 200 for L2 and IS GPS 705 for L5, both included on the DVD in the Misc/directory).

The results of all this activity in the L1 band alone are remarkable. The band is now planned to be full of CDMA signals. This does not include the FDMA signals broadcast by the Russian GLONASS system, which are in the upper part of the band. However, there is a strong possibility that GLONASS will also join the other GNSS systems in providing a CDMA based signal at or near the L1 center of band. Figure 17.2 illustrates just how crowded the L1 band is becoming.

17.2.1 Main Concept of Operation for Galileo

The Galileo satellite constellation has been determined, following a number of studies, to be a Walker (27,3) type with 27 active satellites in three orbital planes inclined at 56° to the equatorial plane with 3 inactive spares. At least one guiding principle was that there should be no common point of failure with any other satellite navigation system. This meant that the constellation would require independent support for all of its functions. This means that the time and geodetic references for Galileo would need to be subject to independent realisation in comparison with

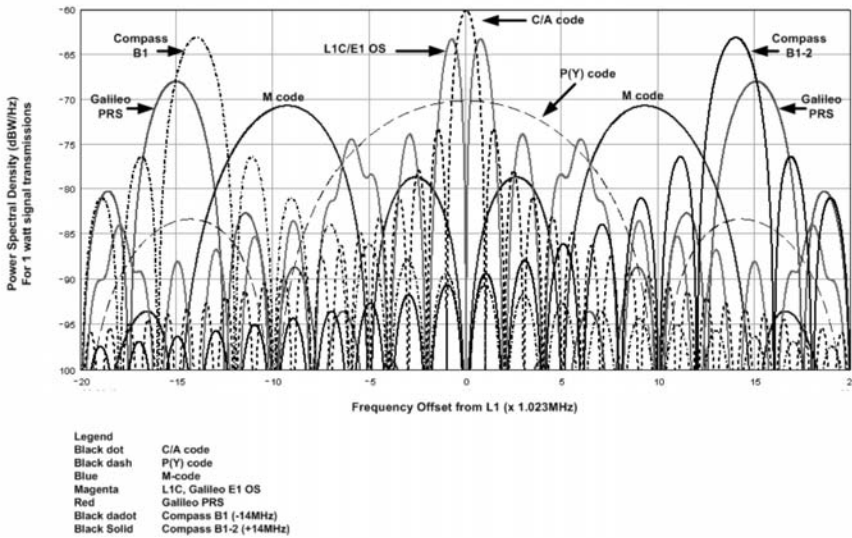


Figure 17.2 CDMA signals in the L1 spectrum. (Color image is included on the DVD.)

either GPS or GLONASS. Such a requirement provides for no common mode of failure with other systems. Such a concept is not fully possible to realize in practice as there are common time standards, such as TAI and UTC, amongst others which support international standards of time interval and time reference. A relationship to common international standards is useful in providing interoperability between systems. Similarly, the International Terrestrial Reference Frame (ITRF) and International Earth Orientation Service (IERS) provide a common underpinning for the basic references for orbital descriptions and the subsequently derived location solutions.

17.3 New Modulation Opportunities

The existing CDMA transmissions from GPS, for C/A code and P(Y) signals, have a rectangular spreading symbol although at different chipping rates, induced by BPSK modulation. The modulation arises from the designation of a constant carrier phase for each code state, or combination of states, to be transmitted. This is satisfactory if there are only a limited number of signals to be used in a signal band. An illustration of this type of modulation is shown in Figure 17.3.

As new requirements to provide satellite navigation services arose, a need was established to have improvements in signal multiplex techniques and in spectrum occupancy. In this section, I will concentrate on the second issue whereby the signals associated with different satellite navigation services could be spread or separated in the frequency spectrum. This has been the prime motivation for the development of different signal types, which are reviewed below. Recognizing that some satellite navigation systems have a variety of user communities—GPS is considered as a dual-use system providing navigation capabilities for both civilian and military users—the option to deny the use of the signals to one community

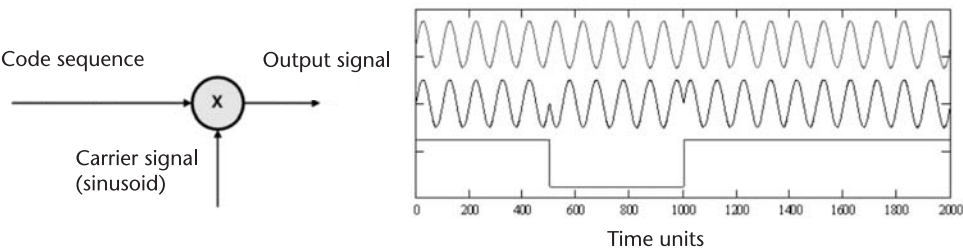


Figure 17.3 BPSK: method and sample waveforms.

while permitting use by the other can be made much more effective with signals that are spectrally separate. To commence this analysis of signal options, the characteristics of the existing rectangular spreading symbols will be briefly explored.

17.3.1 Existing Spreading Symbol—BPSK Modulation

Modulation of the carrier signal is of constant amplitude and at two phases diametrically opposed on a phase diagram. This is illustrated in Figure 17.4. The spectrum of this waveform is the key to further understanding of the new signal evolution. The computation of the spectrum is included in this section in order to illustrate the method that is followed for other spreading symbol types.

The waveform transmitted by the satellite is constructed from a number of components and processes. The first of these is a (spreading) code generator used to uniquely identify the satellite, providing a wideband emission that can be reduced to a response in time of the order of the reciprocal of the bandwidth by suitable signal processing in the receiver. The primary reception technique is a matched filter often implemented by correlation processing either in the time or frequency domains. Data modulation by carrier phase inversion is also applied but at a much lower rate than the code rate. Pilot signals do not have the data modulation component. The composite sum of all the transmitted signals are passed through a high-power amplifier and various, usually distributed, filters. The impulse response of the product of the frequency transfer responses of the filters is convolved with the generated waveform to form the satellites' output signal. These stages are illustrated in Figure 17.5.

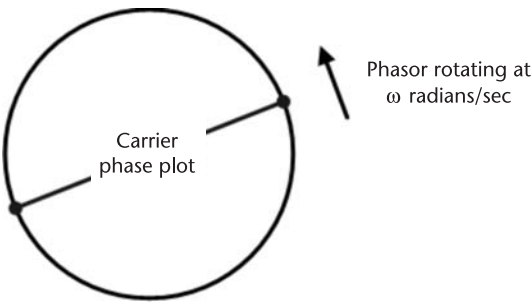


Figure 17.4 Rotating phase plane vector.

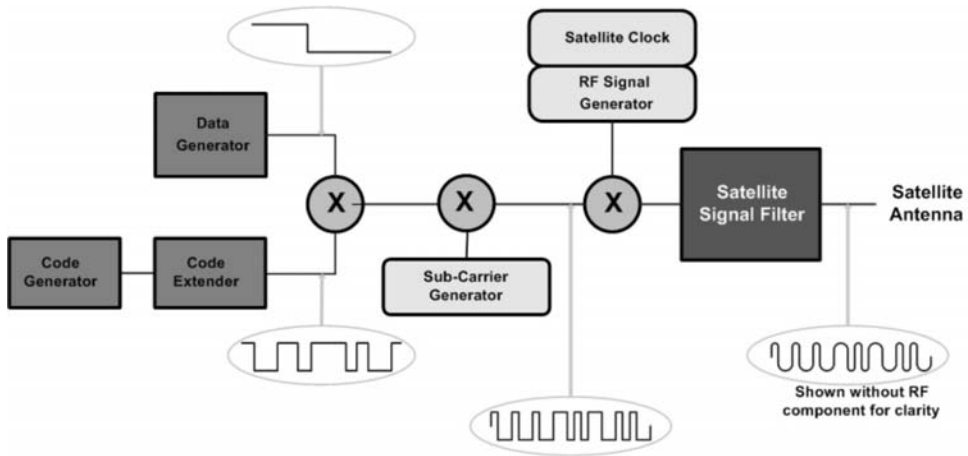


Figure 17.5 Illustration of navigation signal generation in satellite.

The output from the satellite, $r(t)$, results from the convolution of the ‘ideal’ satellite navigation signal, $s_T(t)$ and the (usually distributed) satellite filter $h_T(t)$. This operation is performed in the time domain but the effects of it can often be more easily accounted for in the frequency domain, especially when only the power spectrum is required. The ideal satellite navigation signal is used to phase modulate the carrier and is formed from a code sequence, c_k (at the k th clock pulse), which is extended for the duration of the interval between elements of the sequence. The extended code sequence is modulo-2 added to a navigation data sequence $d(t)$. The extension to the code sequence is of constant amplitude for rectangular BPSK.

In practice the code and data sequences repeat in accordance with the definitions of the system—for GPS C/A code the repetition is every 1 ms for the code and 30 seconds for much of the data content, with a super frame repeating every 12.5 minutes. For the initial analysis of the spectrum of these signals, a simplifying assumption is made that the code and data sequences are of infinite length (that is they do not repeat). This assumption is clearly invalid for the C/A code signal. At the time of writing, this remains the subject of research, though many of the effects are now known. With the assumptions of infinite length, the power spectral density of the “ideal” signal is computed from the correlation function of the code and data sequence. Since the data sequence is considered to be random, this is subsumed into the code sequence. The (auto) correlation function (ACF) of the time-extended code sequence, with a time interval of ΔT between adjacent code elements, is:

$$C(\tau) = \lim_{N\Delta T \rightarrow \infty} \frac{1}{N\Delta T} \left\{ \int_{t \in N\Delta T} c_k s(t - k\Delta T) \cdot c_j s(t - j\Delta T - \tau) \cdot dt \right\} \quad (17.1)$$

where the integral is performed over a duration of N intervals ($N\Delta T$), and then the number of code elements included in the integration is allowed to increase to infinity. The integral gives the following result:

$$C(\tau) = \begin{cases} \frac{1}{\Delta T} \left\{ 1 - \frac{|\tau|}{\Delta T} \right\} & \text{for } |\tau| \leq \Delta T \\ 0 & \text{otherwise} \end{cases} \quad (17.2)$$

This can be viewed as the product of 2 terms (1) the $(1/\Delta T)$ representing the ACF of the code sequence and (2) the triangular function representing the convolution of the two rectangular pulses (the code sequence time extension). Using the Wiener-Khinchine Theorem (a Fourier transform), the power spectral density (PSD), $\Phi(\omega)$, for the ideal rectangular satellite navigation signal is:

$$\Phi(\omega) = \Delta T \cdot \left\{ \frac{\sin(x)}{x} \right\}^2 \quad \text{where } x = 0.5 \cdot \omega \Delta T \quad (17.3a)$$

I will just note that there are two ways of arriving at this result—the direct application of the Wiener-Khinchine Theorem and an alternative (and simpler) method through finding the complex spectrum of the rectangular code extension, using a Fourier transform with support $t \in 0 \dots \Delta T$, and then passing to the PSD by multiplication with the complex conjugate spectrum. Note that the $\sin(x)/x$ function has a nominal singularity at $x = 0$, for which the limiting value is 1, and that the PSD has evenly spaced zeros in ω at:

$$\Phi(\omega) = 0 \quad \text{when } \omega = \frac{2n\pi}{\Delta T} \text{ for all } n \neq 0 \quad (17.3b)$$

A plot of PSD for C/A code and P code signals is given in Figure 17.6. The spectra of these signals are concentrated around the carrier frequency although with different widths resulting from the different spreading code rates. Note that the graphs are drawn for a 1-W signal emission from the satellites (actual P-code signals are typically 3 dB lower than C/A code signals) and with no satellite signal filters. (See Chapter 2 for additional plots of various signal time and frequency characteristics.)

When a filter, with frequency transfer response $H_T(\omega)$, is used to control the output spectrum from the satellite, the new PSD is modified to:

$$\Phi_H(\omega) = \frac{\Phi(\omega)}{\frac{1}{2\pi} \int_{\omega} |H_T(\omega)|^2 \Phi(\omega) d\omega} \quad (17.4)$$

where the denominator re-normalises the integrated power in $\Phi_H(\omega)$ back to unit area (equivalent to 1 watt emission). The curves in Figure 17.6 also show the effects

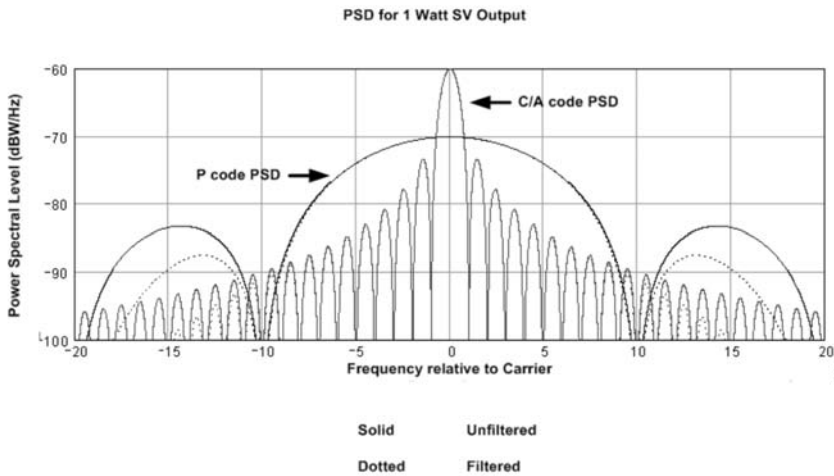


Figure 17.6 PSD for C/A and P-code emissions.

of such a satellite filter with a bandwidth of 30.69 MHz and a frequency transfer function from a fifth-order Butterworth filter—this is the dotted line below the P-code spectrum. The effect of this is insignificant for C/A code signals but results in an amplitude correction of +0.284 dB for the P-code emission.

17.3.2 Binary Offset Carrier (BOC) Modulation

A new form of spreading code definition appears to have resulted from a search for alternatives conducted by John Betz [1] of the Mitre Corporation during the signal definition phase for a new GPS military signal (which resulted in the GPS M-code modulation). A number of alternatives also emerged during related studies by others and, although these have been broadcast by test satellites, they have not yet been adopted in any final signal definition.

17.3.2.1 Concepts and Definition

The simplest description of BOC modulation is through the introduction of a synchronous subcarrier modulation by a binary (square) waveform after the code and/or data modulation stages. This is illustrated in Figure 17.7 and in further detail in Figure 17.8. The subcarrier signal is synchronous with the code generator. BOC modulation is defined by two integers (m , n) defining the code chipping rate and the frequency of the binary subcarrier relative to 1.023 MHz. $\text{BOC}(n, m)$ thus refers to a binary signal with a code chipping rate of $m \times 1.023$ MHz and a binary subcarrier with frequency $n \times 1.023$ MHz. A $\text{BOC}(1, 1)$ spreading symbol has a 1 Mchip/sec code chipping rate and a subcarrier of the same rate so that the spreading symbol has an equal duration at the ‘0’ and ‘1’ logic levels—this is illustrated in Figure 17.8.

From this simple definition, two synchronisation cases can be identified from the phasing of the subcarrier with the code transition times. These are sine and cosine versions:

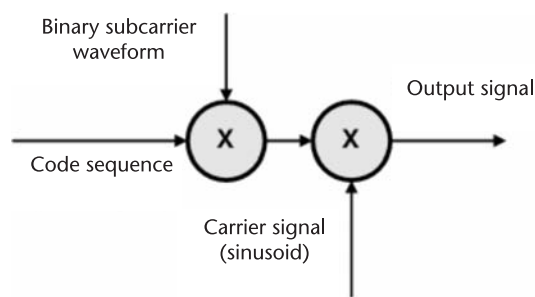


Figure 17.7 Method of binary subcarrier modulation.

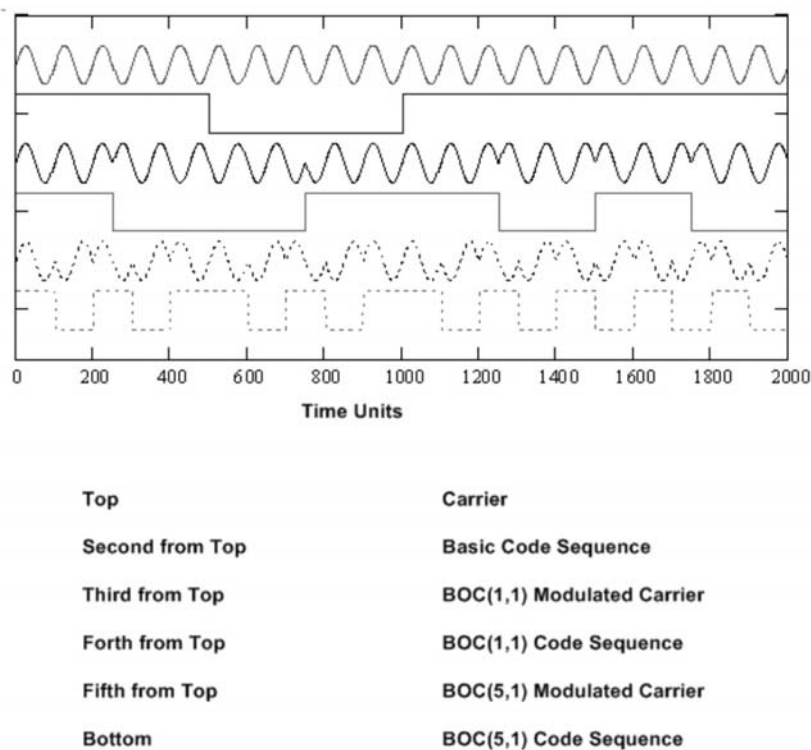


Figure 17.8 Binary subcarrier modulation for BOC(1,1) and BOC(5,1) examples.

$$\begin{aligned} s_{Ts}(t) &= c_k \cdot d(t) \cdot \text{sign}(\sin(n\omega_0 t)) \text{ for sine phasing, and} \\ s_{Tc}(t) &= c_k \cdot d(t) \cdot \text{sign}(\cos(n\omega_0 t)) \text{ for cosine phasing} \end{aligned} \tag{17.5}$$

where $n\omega_0$ is the subcarrier frequency in rad/sec. In principle, there could be a variety of other phasing arrangements but these have not found any practical utility at this stage.

The code chipping rate of $m \times 10.23\text{MHz}$ corresponds to a code element duration of $\Delta T/m$ where $\Delta T = (1.023 \cdot 10^6)$ secs (a C/A code chip duration). Consequently,

the spreading waveform depends upon the ratio of $n:m$. In practice, two separate cases have been identified depending upon the number of half cycles of the subcarrier that are included in the duration of a code chip—the number, p , could be even or odd ($p = 2n/m$). In principle, a wider range of choices is theoretically available with arbitrary n/m values with the definition of a synchronous subcarrier. However, none of these have so far found any practical application, partly due to the fact that the subcarrier synchronization would have to be defined over a multiple number of code elements (chips) up to m . Such cases were studied during the design phase of Galileo but not adopted, primarily as a result of the additional complexity in navigation signal generation.

The four remaining cases (sine or cosine phasing and even/odd number of subcarrier half cycles per code chip), which apply to the envelope of current applications, are illustrated in Figures 17.9 and 17.10. Figure 17.9 contains examples for even-numbered of subcarrier half cycles per chip and Figure 17.10 has examples for the odd-numbered cases. In the next section the PSDs for these four cases will be defined.

17.3.2.2 Spectral Properties

In the previous section, we found two methods for deriving the spectrum of a rectangular spreading symbol convolved with an infinite length ideal random code

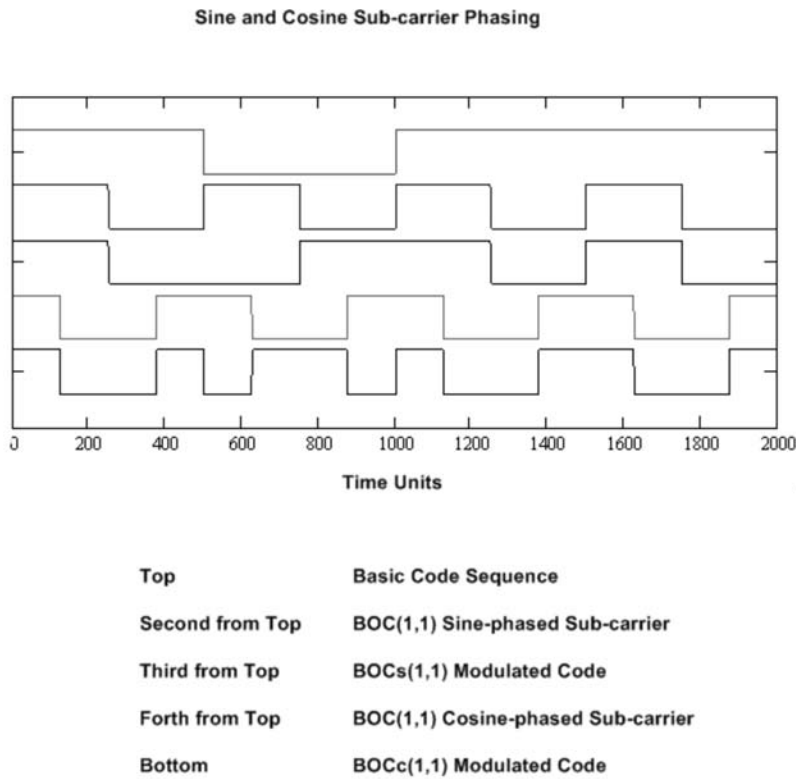


Figure 17.9 Time waveforms for sine and cosine subcarrier phasing (even number of subcarrier half cycles).

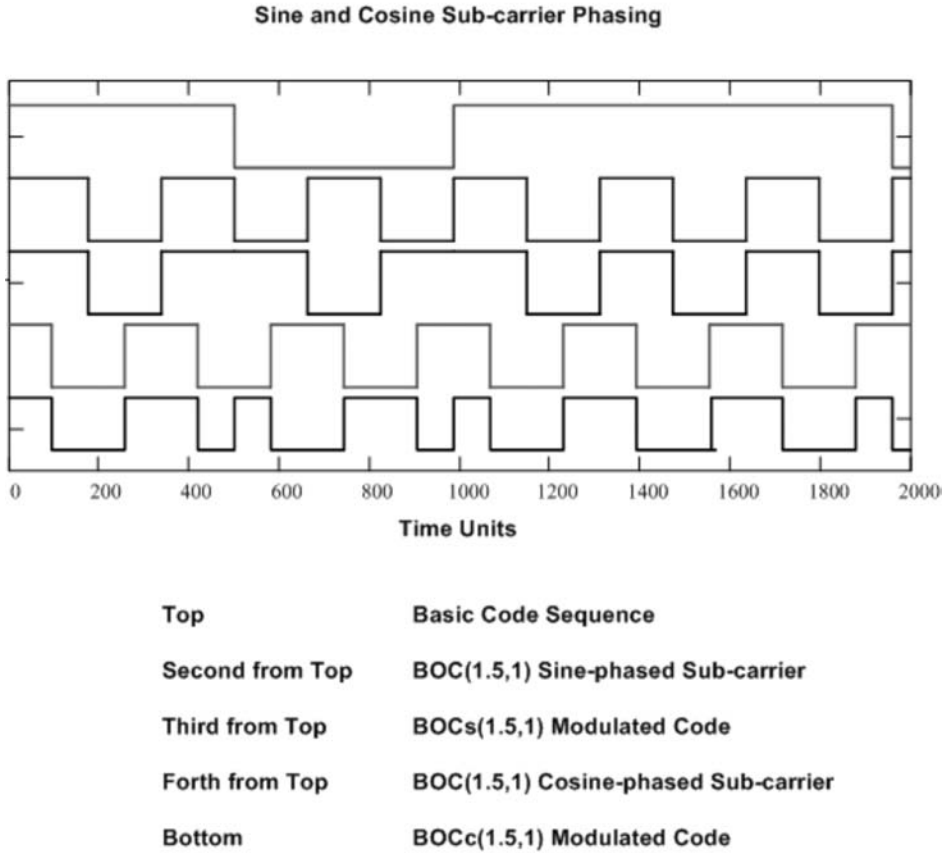


Figure 17.10 Time waveforms for sine and cosine subcarrier phasing (odd number of subcarrier half cycles).

sequence. These methods are general and can be used again to derive the spectral properties of $\text{BOC}(n, m)$ spreading symbol modulation with an ideal code sequence (and, incidentally, with a nonideal finite length repeated code sequence). For the purpose of showing the method, the complex spectrum of the spreading symbol will be derived and converted to the corresponding PSD of the coded navigation signal for the sine phased subcarrier with an even number for ($p = 2n/m$). The details of the derivation are contained in Appendix A, included on the DVD accompanying this book, while the results are presented below. The method can be equally applied to all the other cases but these have not been included for brevity. From the PSD or otherwise, the autocorrelation functions, ACF, can also be derived.

The result for the PSD, as derived in Appendix A, of a sine-phased $\text{BOC}(n, m)$ spreading symbol with $p (= 2n/m)$ even is:

$$\Phi_{S, \text{even}(n, m)}(\omega) = \frac{\Delta T}{m} \cdot \left\{ \frac{\sin\left(\omega \frac{\Delta T}{4m}\right)}{\left(\omega \frac{\Delta T}{4m}\right)} \cdot \frac{\sin\left(\omega \frac{\Delta T}{4n}\right)}{\cos\left(\omega \frac{\Delta T}{4n}\right)} \right\}^2 \quad (17.6)$$

and for cosine phasing for p even, (not derived in Appendix A) is:

$$\Phi_{C_{\text{even}}}(\omega) = \frac{\Delta T}{m} \cdot \left[\frac{\sin\left(\omega \frac{\Delta T}{4m}\right)}{\left(\omega \frac{\Delta T}{4m}\right)} \cdot \frac{\left\{1 - \cos\left(\omega \frac{\Delta T}{4n}\right)\right\}}{\cos\left(\omega \frac{\Delta T}{4n}\right)} \right]^2 \quad (17.7)$$

The expressions for p odd and both sine and cosine subcarrier phasing respectively are:

$$\Phi_{S_{\text{odd}}}(\omega) = \frac{\Delta T}{m} \cdot \left[\frac{\cos\left(\omega \frac{\Delta T}{4m}\right)}{\left(\omega \frac{\Delta T}{4m}\right)} \cdot \frac{\sin\left(\omega \frac{\Delta T}{4n}\right)}{\cos\left(\omega \frac{\Delta T}{4n}\right)} \right]^2 \quad (17.8)$$

$$\Phi_{C_{\text{odd}}}(\omega) = \frac{\Delta T}{m} \cdot \left[\frac{\cos\left(\omega \frac{\Delta T}{4m}\right)}{\left(\omega \frac{\Delta T}{4m}\right)} \cdot \frac{\left\{1 - \cos\left(\omega \frac{\Delta T}{4n}\right)\right\}}{\cos\left(\omega \frac{\Delta T}{4n}\right)} \right]^2 \quad (17.9)$$

The reader should be able to observe a commonality in the structure relating the different PSD expressions. For example, in changing from even to odd numbers of half subcarrier cycles (p), the $\sin(x)/x$ term is changed to a $\cos(x)/x$ term.

The PSDs for three sine-phased BOC PSD, with p even, are plotted in Figure 17.11(a). These provide spectral shapes for known and used signals. The $\text{BOC}_s(1,1)$ was used as a baseline spreading symbol modulation for the early Galileo and GPS III open access signals on L1, although the signal has evolved from this basic PSD by the addition of a second component. $\text{BOC}_s(6,1)$ has been introduced at reduced power levels into the Galileo and GPS L1 open services to provide improved multipath and pseudorange measurement performance. $\text{BOC}_s(10,5)$ is currently in use on L1 and L2 to support M-code signal transmission.

Figure 17.11(b) contains PSDs for cosine-phased spectra and includes known and used spreading symbol modulations including $\text{BOC}_c(10,5)$ used on the Galileo PRS emission in E6, $\text{BOC}_c(15,2.5)$ used for PRS signals in L1 and, for comparison, $\text{BOC}_s(10,5)$ (M-code signals in L1 and L2). Notice the relative side-lobe levels for the sine- and cosine-phased $\text{BOC}(10,5)$ modulation with the cosine-phasing having larger “outer” side-lobes than the “inner” ones and vice versa for sine phased subcarrier modulation. Furthermore, observe that the peak of the mainlobe of the signal energy is slightly lower than the subcarrier frequency for sine-phased modulation and slightly higher for the cosine-phased modulation.

Figure 17.11(c) contains spectral plots for sine- and cosine-phased PSDs with odd numbers of half cycles of the subcarrier. There are not many examples of these in any practical cases with the possible exception of a version of $\text{BOC}_s(15,10)$ used for the phase-coherent transmission of Galileo signals in E5, known as AltBOC. The lower lobe of this transmission corresponds with the QPSK-R10 GPS trans-

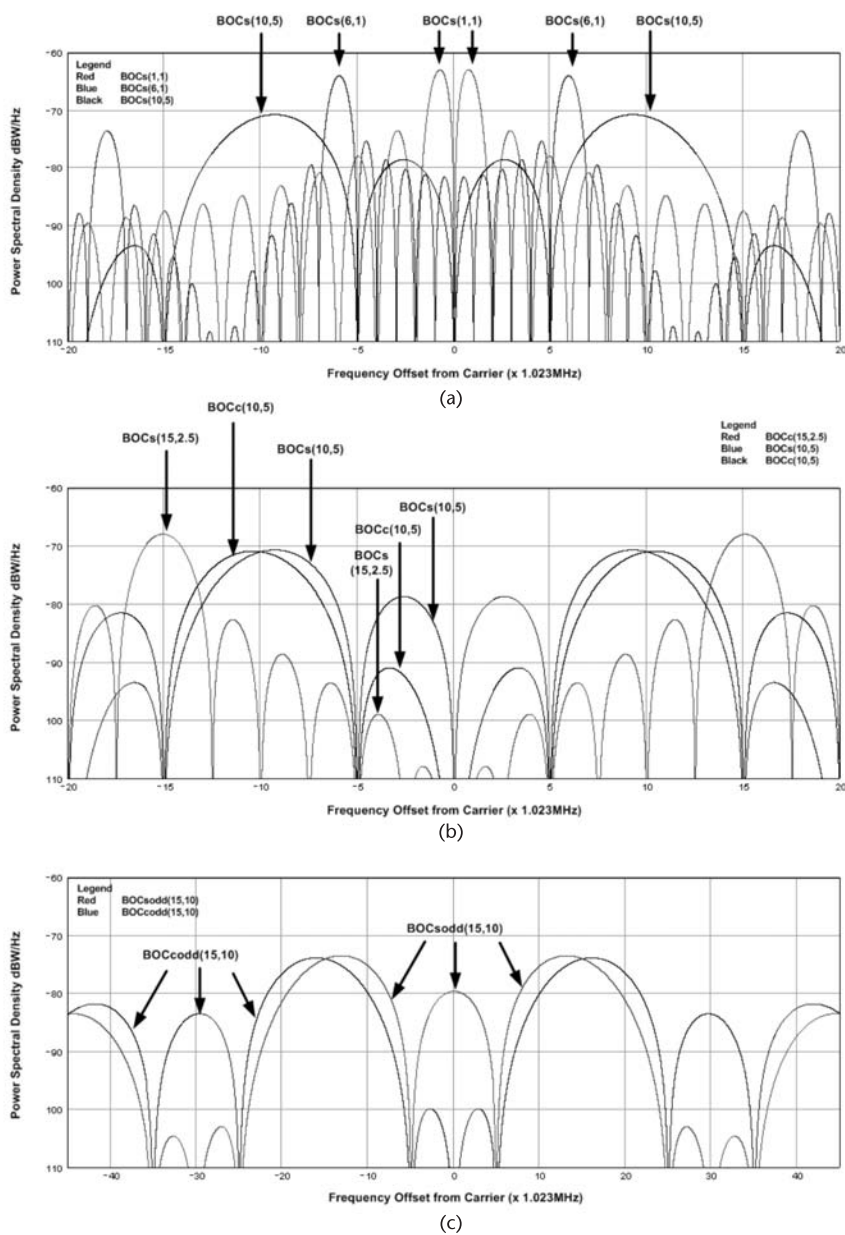


Figure 17.11 (a) Various sine-phased BOC power spectral density forms; (b) various cosine-phased BOC power spectral density forms; and (c) comparison of sine- and cosine-phased (p odd) PSD. (Color image is included on the DVD.)

mission (quadrature phase-shift keying with rectangular spreading symbol at 10 Mchips/sec) in L5 (whose regular transmissions are due to commence with the launch of the Block IIF satellites). The cosine-phase p -odd version has a spectral zero at the subcarrier frequency causing a reduction in PSD values in the region of the transmission frequency.

It should be noted that each of the denominators in the PSD expressions with one or more zeros could lead to potentially singular spectrum values during

evaluation. In practice, the value of the spectrum at these points does not increase without bound—this can be easily determined by taking the algebraic limit values at these points—each of these limit values results in a reasonable and continuous spectrum. I have not included this detail in the chapter but formulae can (and have) been derived to provide values at the limit frequencies.

As a summary, BOC modulation provides for a means to engineer frequency domain separation of signals from different satellite systems or from different services. This allows for the separation of signals primarily used for civil services in modernized satellite navigation systems such as GPS L1C or Galileo L1 open service to be separated from signals intended for military or governmental applications such as GPS M-code or Galileo Public Regulated Service (PRS).

17.3.2.3 Autocorrelation Functions for BOC Modulation

One reason for using BOC spreading symbol modulation is the improvement in measurement accuracy. There are several ways to view this feature. The first of these relates to the Gabor bandwidth of a signal.

It has been widely debated during Galileo signal development that the ability to provide good multipath performance was to increase as far as possible the bandwidth of the transmitted signal. This can be related to the Cramer-Rao lower bound on estimating the location of the peak of a received pulse-type signal. This is essentially the measurement by which the range to each satellite in a visible constellation is estimated. The Cramer-Rao lower bound (CRLB) is a lower bound on the variance of any unbiased (delay) estimator in white Gaussian noise. The CRLB is derived from the conditional probability of the vector of delay measurements given a value for the delay (there are several forms):

$$\sigma_{\tau}^2 \geq \frac{1}{-E \left\{ \frac{\partial^2 [\ln(p(\mathbf{x} | \tau))]}{\partial \tau^2} \right\}} \quad (17.10)$$

After suitable manipulation using the likelihood function for the signal in Gaussian noise [2], this can be shown to be:

$$\sigma_{\tau}^2 \geq \frac{1}{2C} \cdot \frac{1}{\hat{\omega}^2} \quad \text{where } \hat{\omega} \text{ is the rms or Gabor bandwidth} \quad (17.11)$$

The Gabor bandwidth ($\hat{\omega}/2\pi$) is derived from the PSD of the transmitted (or received) signal PSD:

$$\hat{\omega}^2 = \int_{-\infty}^{\infty} \omega^2 \cdot \Phi(\omega) d\omega \quad (17.12)$$

The limits of integration are set at $\pm\infty$ but would, in practice, be determined by the filters inside the transmitting satellite. By maximizing the Gabor bandwidth

for any specific signal, the variance of the measurements is minimized. In general, choosing spreading code modulation with energy peaks farthest away from $\omega = 0$ (relative to the carrier frequency) increases the CRLB. Ultimately, this is limited by the bandwidth of the satellite filters or by the ITU band limit agreements. Figure 17.12 shows how the Gabor bandwidth of some BOC spreading symbols are related to the transmission or reception bandwidth (if the receiver limits the used signal bandwidth, actual measurement variance also increases). These concepts are entirely consistent with the knowledge acquired on multipath mitigation, especially the most modern techniques using, for example, the so-called Vision correlator or those based on Wiener filters [3].

There are assumptions associated with the derivation of the CRLB, which required it to be an unbiased estimator, and some other limitations. Care needs to be taken for BOC spreading symbols to ensure that such assumptions are fully satisfied.

The ACF shows some of the parameters that might lead to difficulty. This leads to a consideration of the form of the ACF for BOC spreading symbols. These can be derived using conventional techniques such as those in Appendix A. To illustrate the phenomenon, I produce the ACF for the cosine-phased BOC(15,2.5) used to support Galileo PRS signals in L1. The infinite bandwidth ACF is shown in Figure 17.13. It demonstrates a multipeak response, with peaks separated by half cycles of the subcarrier modulation. This shows one of the primary concerns in using BOC modulation with large subcarrier offsets ($p = 12$ for PRS, $p = 4$ for M-code). During receiver acquisition, it is possible to commence tracking on an incorrect ACF peak—which leads to large measurement errors (not predicted by the CRLB). The probability of acquisition and tracking on an incorrect peak deteriorates with poorer carrier to noise ratio. A heuristic observation easily provides convincing evidence—the separation of adjacent peaks in magnitude must be larger than the receiver band-limited noise by a factor sufficient for detection (such as 10 dB), otherwise the separate peaks cannot be reliably distinguished.

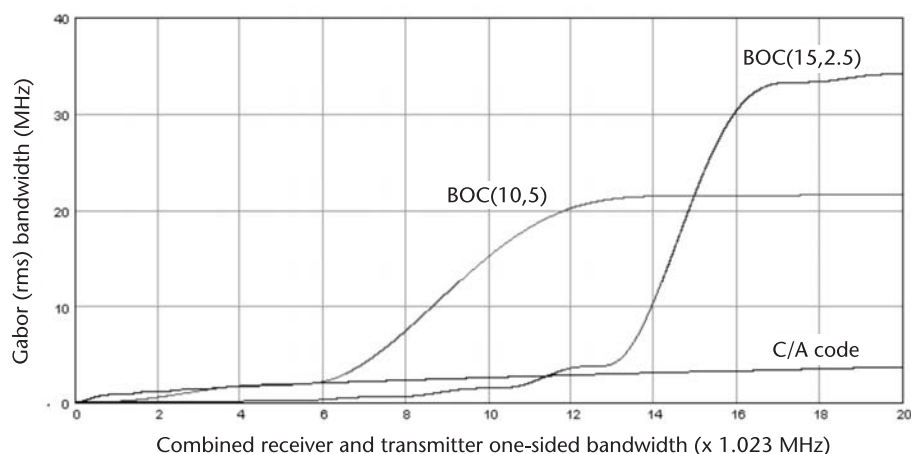


Figure 17.12 Gabor bandwidth for C/A code, M-code, and PRS spectra (rectangular bandpass model).

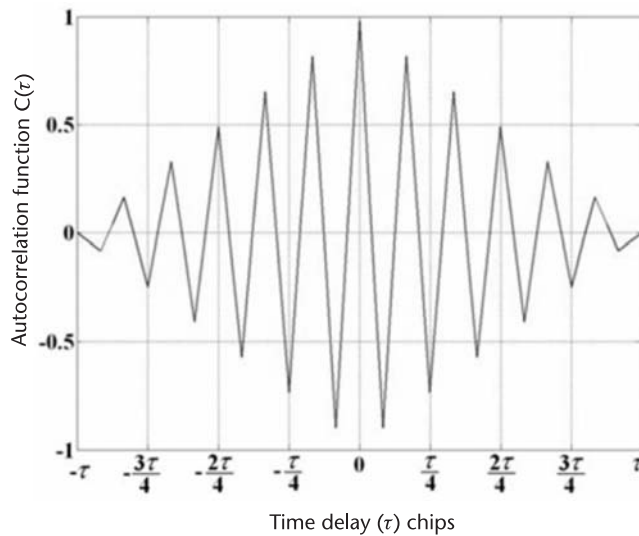


Figure 17.13 Autocorrelation function for BOCc(15,2.5) PRS signal.

Although this seems a depressing verdict on BOC spreading symbols, especially those of high p value, plenty of research effort has been applied to establishing new signal processing arrangements. The key requirement is to find processing that supports low bandwidth measurement or tracking of the subcarrier. The best of the methods found so far provides a means to track the correct subcarrier phase directly. Some of the methods for overcoming this problem are also discussed in Chapter 2 of this book.

17.3.3 Multiplex BOC Modulation

The original E1 OS modulation proposed for Galileo was a BOC(2,2) spreading symbol. The proposal was based on a near two- to three-fold improvement in multipath mitigation and pseudorange measurement variance. However, the difficulty in controlling the wideband spectrum lead to the rejection of this modulation. After a significant search for alternatives in which more control could be exercised on the PSD, a new form of BOC modulation was found in which each spreading design was composed of the combination of two different spreading symbols. There are many variations on this general theme, however, not all of them are useful because of the possible interaction between the different constituent parts of the spreading symbol. The search included several versions of known spreading modulations using tertiary and 8-PSK waveforms—see for example [4]. For additive combinations, there are restrictions placed on the choice of waveforms to avoid measurement bias. One of these is that partner waveforms (spreading symbols) must be anti-symmetric. This restricts the choice to partner with the (baseline) BOC(1,1) modulation to a small number of candidates [5].

One version has been jointly agreed for use by GPS and Galileo at an historic meeting March 9, 2005, for the open service signals in the L1 band [6, 7]. This is known as multiplex BOC (MBOC) and uses a combination of BOC(1,1) and

BOC(6,1) spreading symbol modulation. For the GPS system, this new signal will be designated L1C and transmitted as part of the GPS III upgrade. It is possible that other emerging GNSS providers might also choose to adopt this spreading symbol modulation for their open service signals.

The purpose in developing this form of spreading symbol modulation was to improve the performance of the signals with respect primarily to multipath propagation. For the record, several attempts were made to find combinations of spreading symbol modulations which satisfied a variety of constraints. These will not all be reported here; however, one significant candidate was formed from the combination of a BOC(1,1) basic modulation and a lower power component made from a BCS modulation. BCS modulation is covered briefly below.

There are two forms of spreading symbol modulation that result in the agreed MBOC spectrum. The emissions of MBOC usually employ both a data-bearing carrier and a pilot signal (without data). This permits several realizations of MBOC depending upon the split of power between these two signals. The Galileo system has adopted a signal format using the amplitude addition of the two components (CBOC) whereas GPS has adopted a time multiplex method of combination (TMBOC). The remaining detail in the design of the MBOC signal is the power split between the components. This has been chosen to be 10/11th of the total to the BOC(1,1) signal and 1/11th devoted to the power in the BOC(6,1) component. These two implementation options are described below as the resulting characteristics are not identical. The MBOC power spectral density is, therefore:

$$\Phi_{\text{MBOC}}(\omega) = \frac{10}{11} \Phi_{\text{BOC}(1,1)}(\omega) + \frac{1}{11} \Phi_{\text{BOC}(6,1)}(\omega) \quad (17.13)$$

The MBOC PSD is plotted (with its components) in Figure 17.14(a). Notice the additional lobes at ± 6 MHz and ± 18 MHz. These contribute to multipath mitigation and pseudorange measurement improvements and are available to receivers with sufficient bandwidth; otherwise the performance (for low bandwidth receivers) reverts to that of BOC(1,1) modulation. This can be clearly seen through the variation in Gabor bandwidth with transmission/reception bandwidth in Figure 17.14(b). The MBOC Gabor bandwidth exceeds the BOC(1,1) value beyond 6MHz and a further increase at 18MHz (one sided combined receiver and transmitter rectangular filter). These characteristics are thought to provide for performance growth through improvements in receiver technology or to satisfy user requirements. The Gabor bandwidth for MBOC reaches a factor of x2 greater than C/A code signals at 20MHz (one-sided bandwidth).

One general comment, on these spreading symbols, is that their full versatility has not yet been fully explored. The present versions in use by Galileo and GPS III use the same code sequence for both spreading symbol components. This is not a necessary restriction and it is possible to use separate code sequences for each component. This provides the means to support a further signaling channel either for separate measurement of satellite range (at reduced power level using the secondary component) and/or for data transmission. Data transmission using the secondary channel could be implemented providing the CCF between primary and secondary components is zero. This condition is satisfied for the antisymmetric

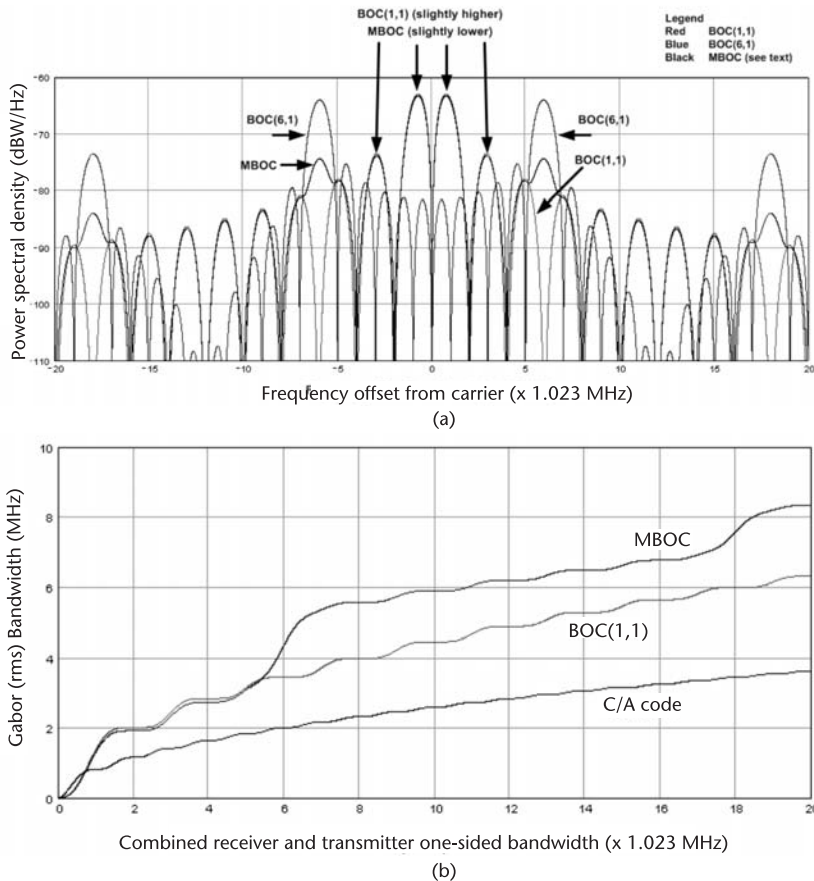


Figure 17.14 (a) MBOC PSD and components and (b) Gabor bandwidth for C/A code, BOC(1,1), and MBOC spectra (rectangular bandpass model). (Color image is included on the DVD.)

sequences used in Galileo CBOC and for the time division multiplex used in GPS L1C.

The reduced power levels available on the secondary component can still provide a suitable data transmission bit error rate (BER) providing signaling at low data rates and/or the use of forward error correction (FEC) are implemented. There could be many uses for such a data channel, including the transmission of time reference data or satellite almanac. If such transmission were confined to the data-bearing channel, the multipath performance benefits available on the pilot channel would not be affected.

17.3.4 Composite BOC Modulation

The composite BOC modulation uses the addition of two signals. The data bearing and pilot signals are formed separately with a common sine-phased BOC(1,1) part and BOC(6,1) parts with opposite signs. Time waveforms for the two spreading symbols are shown in Figure 17.15.

The pilot spreading symbol is chosen to have the largest Gabor bandwidth. The arrangement, with opposite BOC(6,1) phases, is required to cancel unwanted

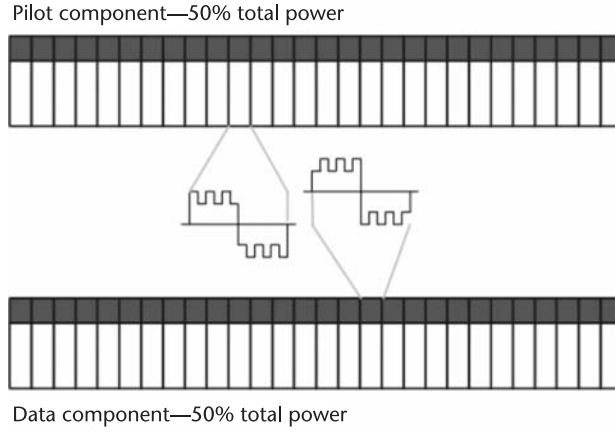


Figure 17.15 Galileo modulation waveforms.

components in the spectrum which otherwise would not satisfy the MBOC spectrum required by the agreement with GPS. The restriction in the formation of the signal is that the pilot and data bearing signals must be equal in magnitude as can be seen in the following equations, where the $S_T(\cdot)$ represent complex spectra (not PSDs):

$$\begin{aligned} S_{\text{TCBOC}}^{\text{Pilot}}(\omega) &= \alpha S_{\text{TBOCs}(1,1)}(\omega) + \beta S_{\text{TBOCs}(6,1)}(\omega) \\ S_{\text{TCBOC}}^{\text{Data}}(\omega) &= d_k \cdot (\alpha S_{\text{TBOCs}(1,1)}(\omega) - \beta S_{\text{TBOCs}(6,1)}(\omega)) \end{aligned} \quad (17.14)$$

The data component is d_k , taking values of ± 1 . The power spectrum for the pilot signal becomes (see Appendix A):

$$\begin{aligned} |S_{\text{TCBOC}}^{\text{Pilot}}|^2 &= |jx^{p/2}|^2 \cdot (\alpha \hat{S}_{\text{TBOCs}(1,1)}(\omega) + \beta \hat{S}_{\text{TBOCs}(6,1)}(\omega))^2 \\ \text{where } S_T(\cdot) &= \hat{S}_T(\cdot) \cdot \left(je^{j\omega \frac{\Delta T}{2m}} \right) \end{aligned} \quad (17.15)$$

Notice that the $|\cdot|^2$ component in the right hand side has unity magnitude and is only dependent on the spreading symbol length ($\Delta T/m$), and is identical for BOC(1,1) and BOC(6,1) modulation. It can therefore be factored out of the equation. The PSD for the pilot signal is then:

$$\begin{aligned} \Phi_{\text{CBOC}}^{\text{Pilot}}(\omega) &= \alpha^2 \Phi_{\text{BOC}(1,1)}(\omega) + \beta^2 \Phi_{\text{BOC}(6,1)}(\omega) \\ &\quad + 2 \frac{m}{\Delta T} \alpha \beta \hat{S}_{\text{TBOCs}(1,1)}(\omega) \hat{S}_{\text{TBOCs}(6,1)}(\omega) \end{aligned} \quad (17.16)$$

Note that the resulting spectra after the removal of the $|\cdot|^2$ are real. In this case, the pilot PSD does not conform to the specified MBOC PSD due to the

addition of the cross-spectra terms. These are removed by a data-bearing signal providing the power in each component is identical:

$$\begin{aligned}\Phi_{\text{CBOC}}^{\text{Data}}(\omega) &= \alpha^2 \Phi_{\text{BOC}(1,1)}(\omega) + \beta^2 \Phi_{\text{BOC}(6,1)}(\omega) \\ &\quad - 2 \frac{m}{\Delta T} \alpha \beta \hat{S}_{\text{TBOCs}(1,1)}(\omega) \hat{S}_{\text{TBOCs}(6,1)}(\omega) \quad (17.17) \\ \text{where } S_T(\cdot) &= \hat{S}_T(\cdot) \cdot \left(j \cdot d_k \cdot e^{j\omega \frac{\Delta T}{2m}} \right)\end{aligned}$$

Note also that the pilot and data bearing PSDs are different. They are chosen so that the pilot spectrum has the largest high-frequency component, thereby maximizing the multipath mitigation and measurement performance. The transmitted PSD is, therefore:

$$\begin{aligned}\Phi_{\text{CBOC}}(\omega) &= 0.5 \Phi_{\text{CBOC}}^{\text{Pilot}}(\omega) + 0.5 \Phi_{\text{CBOC}}^{\text{Data}}(\omega) \\ &= \alpha^2 \Phi_{\text{BOC}(1,1)}(\omega) + \beta^2 \Phi_{\text{BOC}(6,1)}(\omega) \quad (17.18) \\ \text{where } \alpha &= \sqrt{\frac{10}{11}} \quad \beta = \sqrt{\frac{1}{11}}\end{aligned}$$

No other combined solution (composite BOC modulation) can satisfy the MBOC PSD requirement. This restriction does not, however, apply to the TMBOC modulation method described in the next section nor to a CBOC method that spreads the signal definition over two code chips. In this case, the sign of the BOC(6,1) component on the pilot and data-bearing components can be alternated from chip to chip. This ensures that the spectra of the individual pilot and data-bearing components each meets the MBOC requirements. The power split between pilot- and data-bearing signals may then be chosen arbitrarily. This option was studied for Galileo E1 OS but rejected due to issues of signal generation in the satellites.

17.3.5 Time Multiplex BOC Modulation

Time multiplex BOC (TMBOC) has been selected for GPS L1C spreading symbol modulation. As the name suggests, the spreading symbol transmitted is either a BOC(1,1) or BOC(6,1) component, both sine-phased. Since the spreading symbols have a time duration of exactly 1 chip, the ideal random properties of the code sequence provide a correlation value of 0 between adjacent code elements. Consequently, there are no cross-spectral components as are present in CBOC modulation. This arrangement permits unrestricted choice of power levels for the data-bearing and pilot components and the possibility to select the proportion of total available BOC(6,1) power that is provided to the pilot component. GPS L1C has published, in IS GPS 800, proposed signal characteristics which assign 25% of the prefiltered transmit power to a data-bearing component, which uses only BOC(1,1) spreading

symbols. The pilot component, providing the best signal for coherent integration, multipath mitigation and measurement performance, will comprise 75% of the transmitted power and use a power division of (29/33) to the BOC(1,1) component and (4/33) to the BOC(6,1) component. This is summarized in the following equations:

$$\begin{aligned}\Phi_{\text{MBOC}}(\omega) &= 0.25\Phi^{\text{Data}}(\omega) + 0.75\Phi^{\text{Pilot}}(\omega) \\ \Phi^{\text{Data}}(\omega) &= \Phi_{\text{BOC}(1,1)}(\omega) \\ \Phi^{\text{Pilot}}(\omega) &= \left(\frac{29}{33}\right)\Phi_{\text{BOC}(1,1)}(\omega) + \left(\frac{4}{33}\right)\Phi_{\text{BOC}(6,1)}(\omega)\end{aligned}\quad (17.19)$$

where $\frac{1}{4} + \frac{3}{4} \cdot \frac{29}{33} = \frac{10}{11}$ as required.

Clearly, many different divisions of power between the data and pilot components are possible while still satisfying the MBOC PSD requirement. These were considered but not selected for application. A diagram showing the arrangement of the different TMBOC spreading symbols is provided in Figure 17.16.

A further parameter selection for TMBOC modulation is the choice of the location for the four BOC(6,1) components in the pilot sequence. In principle, these could be located anywhere in a repeating sequence of 33 chips; for example, in the first four locations. However, a receiver channel adapted to receive the BOC(1,1) component, receives only noise during the BOC(6,1) emission (due to the feature of antisymmetric sequences). Reception during these intervals can be set to a gain of zero, thereby reducing the effective noise level.

It also emerged during development that the excising (removal) of certain code elements can improve the cross-correlation performance between different code sequences by up to 1 dB. This is an advantage in intrasystem noise and interference and during reception in weak signal conditions. The locations of the BOC(6,1) spreading symbols are still subject to further optimization with the motivation to

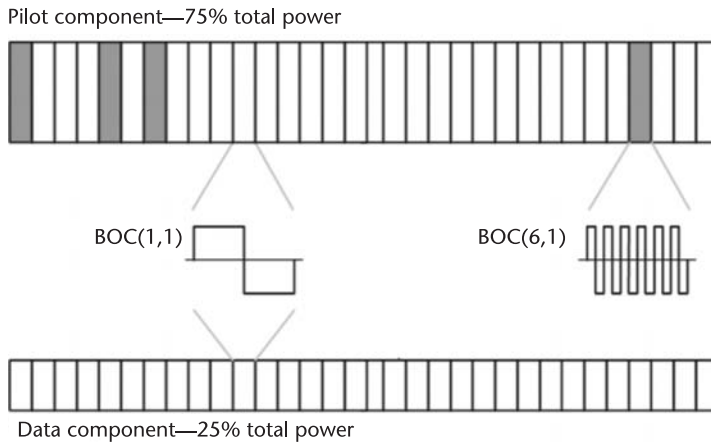


Figure 17.16 GPS L1C modulation waveforms.

obtain better cross-correlation performance. Such signal excision can also be applied to the Galileo CBOC signal but results in a loss of SNR (rather than the opposite for GSP L1C). Nevertheless, the concept of the excision of certain code elements to provide improvement in code performance is an exciting aspect that will attract future research interest.

17.3.6 Other Spreading Symbol Modulation Options

Many different spreading modulation options have been investigated by the EU Signal Task Force either directly or indirectly. Many options were rejected after examination. A particularly comprehensive study is to be found in [8]—a Ph.D. thesis whose study topic was directed at GNSS signal choices, and which contains an extensive bibliography for further reference.

A list of useful spreading symbol modulations, which might have some utility in future systems but have been rejected for application in current systems evolution, would include the following:

1. Linear offset carrier (LOC), sometimes known as Sinusoidal offset carrier;
2. Binary coded symbol (BCS);
3. Double BOC.

A number of other cases can be found in [8].

LOC modulation was implemented in the first Galileo test satellite, Giove A. The binary subcarrier modulation is replaced with a linear modulator using a sinusoidal subcarrier. The resulting spectrum (PSD) can be easily computed and lacks the higher order harmonics of the binary subcarrier present in BOC modulation. Some of these are removed by filters in the satellite, but closer-in harmonics of the spreading code are at lower levels in LOC compared with BOC modulation.

There are two main disadvantages to LOC modulation. First, the satellite high-power amplifier (HPA) cannot be driven into compression in order to avoid amplitude-to-phase conversion products. This reduces navigation signal output power. Second, the reduced level of high-frequency components degrades the range estimation accuracy and the performance of multipath mitigation techniques—especially the more advanced ones. LOC(1,1) and BOC(1,1) power spectral densities are illustrated in Figure 17.17, showing the significantly lower levels of higher frequency component in the former.

BCS first came to the attention of the E.U. Signal Task Force through a note produced by Hegarty and Betz of the Mitre Corporation. It was extensively examined, and from it a number of possible composite signal formats were proposed. One composite added a BCS signal options to a GPS/Galileo BOC(1,1) baseline signal. Various BCS patterns were studied in detail; one of these became a candidate component for Galileo E1 OS evolution in additive combination with a BOC(1,1) primary component. This has been extensively reported in the literature [8].

The basic concept of BCS is that each spreading symbol is constructed from a binary code where each BCS symbol lasts for a specific fraction of the code element duration. This is illustrated in Figure 17.18 for a 10-segment sequence. Other subdivision arrangements are possible, limited by the growth in spectrum width.

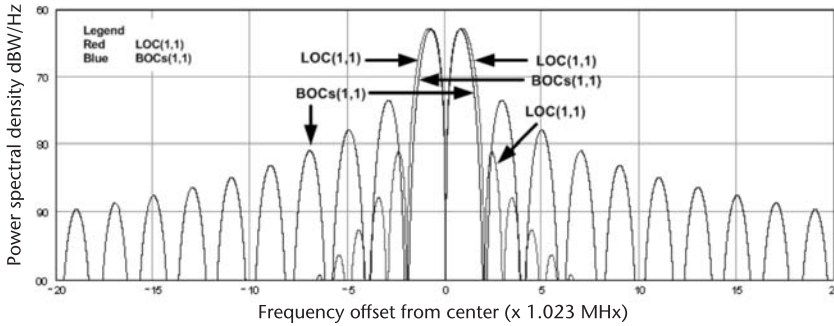


Figure 17.17 LOC versus BOC spectra. (Color image is included on the DVD.)

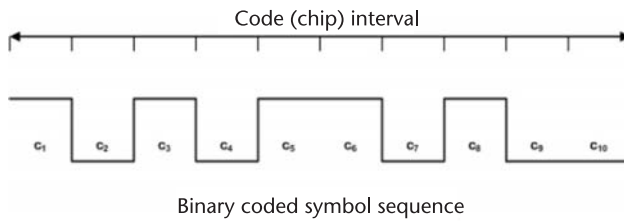


Figure 17.18 Ten-element binary coded spreading symbol.

The individual symbol sequence, $\{c_k\}$, may be chosen in any way but some choices are better than others from a performance viewpoint. The performance metrics are to be determined by the user—for example, one might be to achieve a close approximation to a required spectral shape. For Galileo, the performance metrics concentrated on multipath sensitivity—the area enclosed within the multipath envelope.

For Galileo, a search was conducted of all possible sequences of BCS for up to at least 12 equal segments of the code duration (chip). The best candidates were subject to further analysis before a final version was chosen. The sequence was close to a regular binary modulation, BOC(5,1). As is now known, this is not an antisymmetric sequence and consequently is responsible for measurement bias in a GNSS receiver when used in conjunction with a BOC(1,1) primary component. The measurement bias may be overcome through the process of alternating the BCS sequence (alternately multiplying the sequence by +1 and −1, thereby spreading the BCS sequence over two chips). Some of the sequences correspond to “regular” BOC modulation whereby:

$$c_k = (-1)^k \quad \text{for } k \in 0 \dots (2n - 1); \quad \text{integer } n \text{ corresponding to BOC}(nm, m) \text{ modulation} \quad (17.20)$$

During the Galileo search, BOC(5,1) and BOC(6,1) were both identified as possible candidate BCS sequences but did not perform as well as the chosen one.

This was $\{1, -1, 1, -1, 1, -1, 1, -1, 1\}$. This sequence is nearly a BOC(5,1) with the exception of the last symbol (+1 instead of -1).

During the development of composite modulation sequences, a version of BOC(1,1) was additively combined with this sequence. The power split was 80% to BOC(1,1) and 20% to the BCS sequence. This provided nearly 40% improvement in multipath sensitivity (reduction).

One intriguing concept in the use of BCS is the possibility that a (short) random code could be transmitted as modulation of the primary code element. Although the subdivision of the chip intervals into segments could only provide a short sequence, the resulting spectrum could be made relatively uniform. This might imply a means to minimise interference with other users. However, it has since emerged that this is the poorest choice (see Section 17.5).

Double binary offset carrier modulation is, as its name implies, a modulation of the basic rectangular code chip waveform by two subcarriers. This modulation provides a further splitting of the spectrum in the regions of the peak energy locations of a BOC modulated spreading symbol. An illustration of the spreading symbol modulation is provided in Figure 17.19 for subcarriers at 10 Mc/s and 1 Mc/s based in a 1M chip per second code rate.

In principle, such spreading symbol modulations can be tracked using separate and synchronized subcarrier generators. The arrangement would ensure frequency and phase synchronization between code sequence and both subcarrier generators.

Clearly many similar arrangements can be found using a combination of multiple subcarriers and binary coded sequences. However, it seems unlikely that many of these offer any significant performance advantages over the simpler schemes.

17.3.7 Alternative BOC (AltBOC) Modulation

The Alternative BOC (AltBOC) modulation has been adopted for Galileo application in the E5 band. It is conceptually similar to BOC modulation but with an important difference. The simplest means of description is that AltBOC is a digital quadrature *independent-sideband* modulation technique. BOC modulation does not provide independent sidebands. AltBOC, thus, provides four independent carriers carrying (potentially) four spreading codes. The carriers can be data-bearing

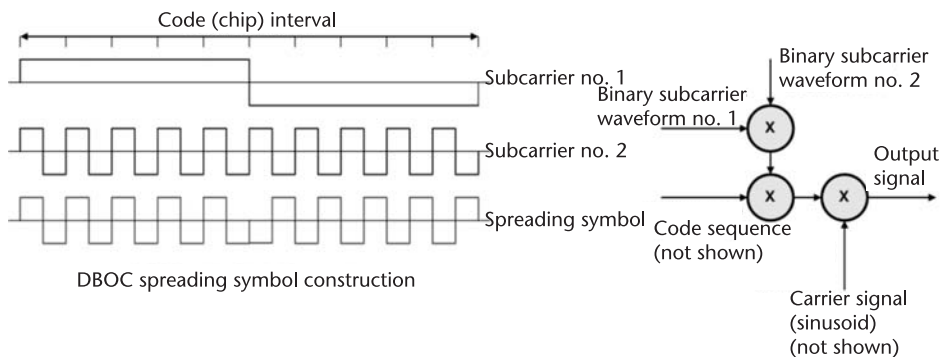


Figure 17.19 Illustration of double BOC signal modulation.

or dataless (pilot) channels. For Galileo, there are two data-bearing signals and two pilot signals, each have different code sequences with low cross-correlation. In Galileo, the carrier offset is 15×1.023 MHz and the code chipping rate 10×1.023 MHz. The carrier frequencies are 1176.45 MHz and 1207.14 MHz, corresponding to 115th and 118th harmonic of 10.23 MHz. The lower BOC sideband corresponds with the GPS L5 signals—which also have a chipping rate of 10.23 MHz (10 Mc/s).

The AltBOC signal is fully coherent with all carriers and codes in synchronization. It can, therefore, be received in a single receiver (AltBOC) channel or separately in a lower and upper frequency channel (as can all BOC signals). In its single channel reception form, the signal offers (for Galileo) a bandwidth of over 90 MHz (limited by the satellite transmission filter) and unprecedented range measurement accuracy. Early measurements from the Giove A test satellite by the European Space Agency (ESA) have shown low susceptibility to multipath signal conditions—much better than any other satellite navigation signal to date. It is expected to offer superlative performance for land survey applications.

Similar to the BOC modulation, and for simplicity, AltBOC is generally referred to as AltBOC(n, m), where ‘ n ’ is the subcarrier offset frequency and ‘ m ’ is the code frequency rate (in multiples of 1.023 MHz). Galileo uses AltBOC(15,10).

AltBOC modulation uses a *complex* subcarrier so that the spectrum is not split up, as is the case of BOC, but simply shifted to higher or lower frequencies. One of the original public disclosures on AltBOC is found in [9] where the signal is defined as the product of a PRN code sequence with a complex subcarrier.

The AltBOC signal can be formed by two (only data signals) or four codes (data and pilot). For the case of two codes, the signal is can be expressed as follows:

$$\begin{aligned} s_{\text{AltBOC}}(t) &= c_{L,k} \cdot c_S(t) + c_{U,k} \cdot c_S^*(t) \\ c_S(t) &= \text{sign}[\cos(\omega_s t)] + j \text{sign}[\sin(\omega_s t)] \\ &= c_r(t) + j s_r(t) \end{aligned} \quad (17.21)$$

where the c_L and c_U terms are the code values in the k th interval (containing the time t). These are modulated with complex subcarriers, c_S and c_S^* . The effect of multiplication with a complex sinusoid (and its conjugate, c_S^*) is to change the frequency either up or down by the related frequency term (ω_s). This is a means of dual digital frequency translation of two independent code sequences producing an upper and lower digital BPSK modulation. The next step in signal definition is the addition of two further codes, modulated with data components in phase quadrature to the dataless codes above:

$$\begin{aligned} s_{\text{AltBOC}}(t) &= (d_{L,k} c_{L1,k} + j \cdot c_{L2,k}) \cdot c_S(t) \\ &\quad + (d_{U,k} \cdot c_{U1,k} + j \cdot c_{U2,k}) \cdot c_S^*(t) \\ c_S(t) &= \text{sign}[\cos(\omega_s t)] + j \text{sign}[\sin(\omega_s t)] \end{aligned} \quad (17.22)$$

The addition of d_L and d_U provides for data modulation on the lower and upper in-phase channels, modulated by codes c_{L1} and c_{U1} during the time interval,

k , as before. In the lower and upper quadrature channels the code sequences are c_{L2} and c_{U2} with no data component.

Unfortunately, the expression above does not have a constant envelope and has nine phase points in a phase plot—eight of these are equally spaced around the constant amplitude circle and one is the central zero point. Such a signal is likely to suffer significant distortion in satellite high-power amplifiers.

A further modification to the AltBOC modulation system (known formally as modified AltBOC) adds an additional signal component (an intermodulation term), which distributes the zero amplitude states to the other (constant amplitude) eight phase states. The intermodulation term does not contain usable information and is removed by the process of signal correlation upon reception. It does, however, contain transmitted energy that reduces the efficiency of the modulation scheme. Signal multiplex schemes (apart from QPSK) all suffer from a similar condition and so this loss is normally acceptable (it is used as a metric to select the best multiplex schemes).

$$\begin{aligned}
 s_{m\text{-AltBOC}}(t) = & (d_L c_{L1} + jc_{L2}) \cdot \left\{ sc_d(t) - jsc_d\left(t - \frac{\Delta T}{4n}\right) \right\} \\
 & + (d_U c_{U1} + jc_{U2}) \cdot \left\{ sc_d(t) + jsc_d\left(t - \frac{\Delta T}{4n}\right) \right\} \\
 & + d_U c_{U1} c_{U2} \cdot (c_{L2} + jd_L c_{L1}) \cdot \left\{ sc_p(t) - jsc_p\left(t - \frac{\Delta T}{4n}\right) \right\} \\
 & + d_L c_{L1} c_{L2} \cdot (c_{U2} + jd_U c_{U1}) \cdot \left\{ sc_p(t) + jsc_p\left(t - \frac{\Delta T}{4n}\right) \right\}
 \end{aligned} \tag{17.23}$$

The changes in the subcarrier waveforms result in the following definitions:

$$\begin{aligned}
 sc_d(t) = & \left\{ \frac{\sqrt{2}}{4} \text{sign} \left[\cos \left(\omega_s t - \frac{\pi}{4} \right) \right] + \frac{1}{2} \text{sign} [\cos(\omega_s t)] \right. \\
 & \left. + \frac{\sqrt{2}}{4} \text{sign} \left[\cos \left(\omega_s t + \frac{\pi}{4} \right) \right] \right\} \\
 sc_p(t) = & \left\{ -\frac{\sqrt{2}}{4} \text{sign} \left[\cos \left(\omega_s t - \frac{\pi}{4} \right) \right] + \frac{1}{2} \text{sign} [\cos(\omega_s t)] \right. \\
 & \left. - \frac{\sqrt{2}}{4} \text{sign} \left[\cos \left(\omega_s t + \frac{\pi}{4} \right) \right] \right\}
 \end{aligned} \tag{17.24}$$

From these equations, the two subcarrier waveforms can be plotted. These are no longer binary as in the original AltBOC definition and are shown in Figure 17.20. However, these do have the advantage that their modulus is always unity:

$$|sc|^2 = sc_d^2(t) + sc_p^2(t) \equiv 1 \quad \text{for all } t \tag{17.25}$$

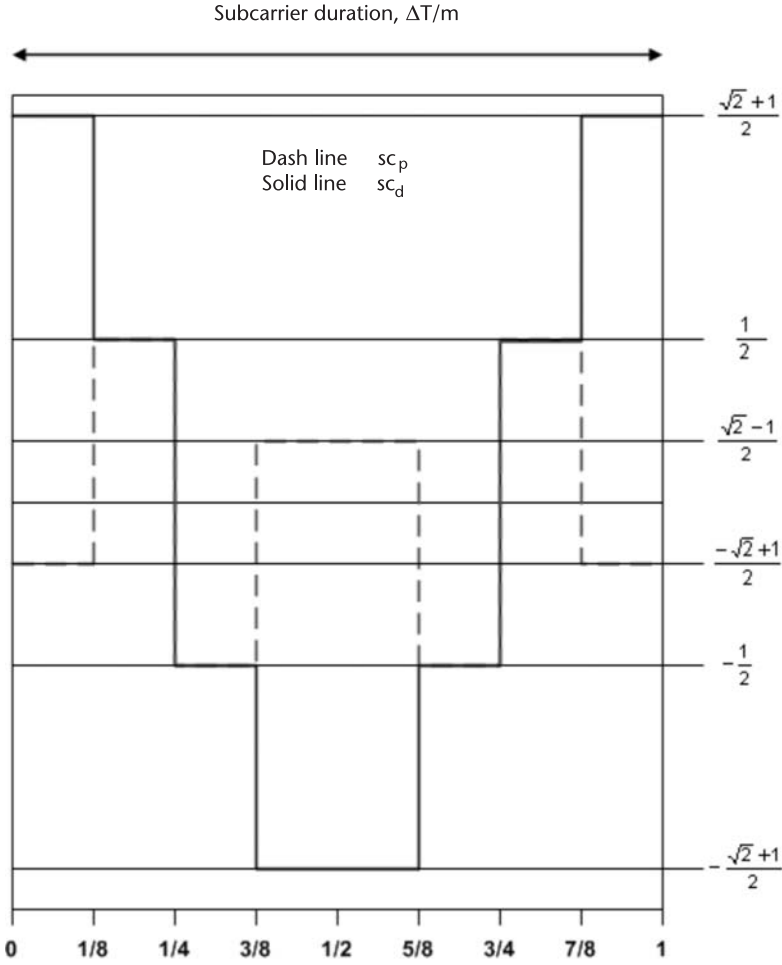


Figure 17.20 Waveform for modified AltBOC subcarriers.

The PSD for AltBOC are not derived here but can be found using the techniques in Appendix A. There are two PSDs depending upon whether there is an odd or even number of subcarrier half cycles (p) in each code chip. Galileo uses p odd. The PSD's are:

$$\begin{aligned}
 \Phi_{AltBOC}^{even}(n, m)(\omega) = & \frac{16m}{\omega^2 \Delta T} \left\{ \frac{\sin\left(\omega \frac{\Delta T}{2m}\right)}{\cos\left(\omega \frac{\Delta T}{4n}\right)} \right\}^2 \\
 & \cdot \left\{ \cos^2\left(\omega \frac{\Delta T}{4n}\right) - \cos\left(\omega \frac{\Delta T}{4n}\right) \right. \\
 & \left. - \cos\left(\omega \frac{\Delta T}{4n}\right) \cos\left(\omega \frac{\Delta T}{8n}\right) + 2 \right\} \quad (17.26a)
 \end{aligned}$$

$$\Phi_{AltBOC}^{\text{odd}}(n, m)(\omega) = \frac{16m}{\omega^2 \Delta T} \left\{ \frac{\cos\left(\omega \frac{\Delta T}{2m}\right)}{\cos\left(\omega \frac{\Delta T}{4n}\right)} \right\}^2 \cdot \left\{ \cos^2\left(\omega \frac{\Delta T}{4n}\right) - \cos\left(\omega \frac{\Delta T}{4n}\right) - \cos\left(\omega \frac{\Delta T}{4n}\right) \cos\left(\omega \frac{\Delta T}{8n}\right) + 2 \right\} \quad (17.26b)$$

The PSD for p odd AltBOC(15,10) and p even AltBOC(15,7.5) are provided in Figure 17.21. The p even AltBOC is provided for reference—it is not used in any known system at this time.

Even though modified AltBOC has a uniform magnitude, there may still be problems in the practical aspects of amplification and transmission. These relate to the usual requirement to distribute the filtering in the satellite transmitter path. As a result, the waveform appearing at the input to the HPA is of limited bandwidth and is no longer of constant magnitude. The precise excursions of the signal are of continuing research interest in improving the performance of actual AltBOC waveforms.

17.4 Signal Multiplex Techniques

Even the simplest GNSS transmits several signals on one carrier. Early GPS signals used a simple system based on the modulation of phase-quadrature components of the carrier to carry two signals—P-code and C/A code. As satellite navigation systems have evolved, the need arised to carry larger numbers of signals (and/or servcies). An example is in the current GPS Block II R-M satellites, which carry P(Y) code, M-code and C/A code signals on the L1 carrier. Galileo test satellites

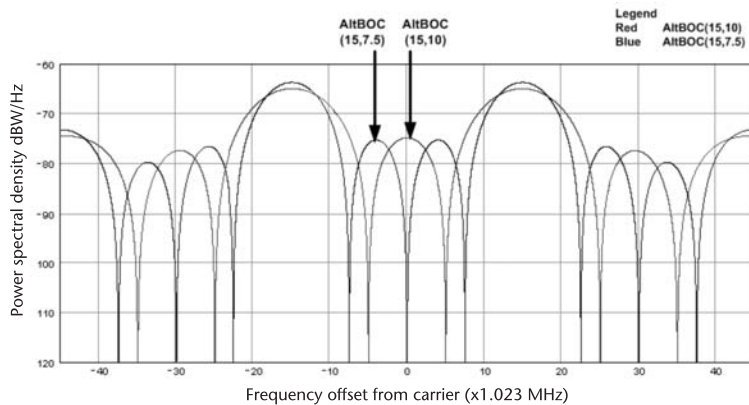


Figure 17.21 AltBOC spectra for Galileo E% and AltBOC_{even}. Color image is included on the DVD.

also transmit, for example, two open service signals (data-bearing and a pilot) as well as a PRS signal on the E1 carrier (1575.42 MHz). In all these cases, there is a requirement to maximise signal power output. This usually means driving the HPA into amplitude compression with the consequent increase in unwanted signal components. This has led to the idea of maintaining the magnitude of the navigation signal constant at all times.

With this as a preface, it is easy to understand the design objectives in seeking signal multipath techniques that support a greater number signals on the same carrier. A variety of techniques have been studied. Galileo has now deployed the AltBOC modulation scheme, which multiplexes four independent signals onto the same carrier. As this also involves the introduction of an unwanted (but not useful signal component), it should be considered as a signal multiplex method. The main methods are briefly identified below.

17.4.1 QPSK

QPSK is the simplest of all the multiplex schemes and does not involve any losses. Each of the two required signal components is placed on one of two orthogonal carrier phases (in-phase and quadrature phase). The signal components can be independent adjusted in amplitude, forming (for binary signals) four points on a carrier phase plane plot. This is illustrated in Figure 17.22. The best known of transmitted signals using this multiplex arrangement is the pre M-code GPS satellites on the L1 frequency. The L5 transmissions from GPS will also use this method.

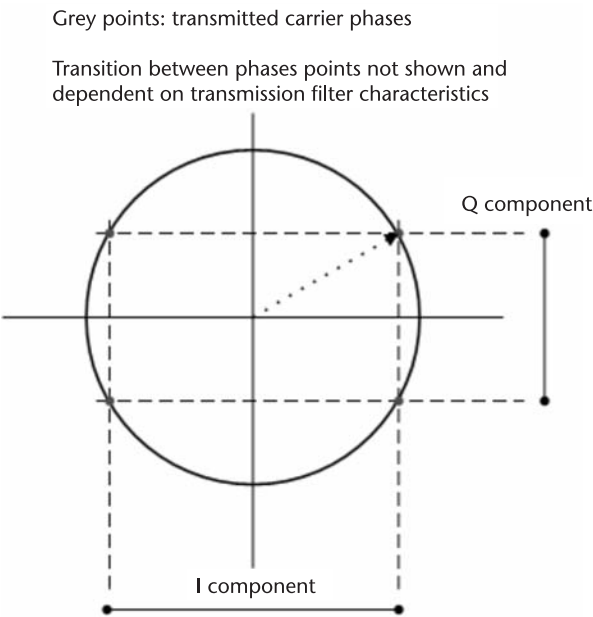


Figure 17.22 Phase plane plot for QPSK multiplex.

17.4.2 Interplex

Interplex modulation is used to multiplex more than two signal components on to a single carrier. In its simplest form, three components can be combined to form the required transmission resulting in a nonconstant signal magnitude. This arrangement has become known as *tri-code hexaphase* modulation. However, the addition of a fourth non-useful component can ensure that the magnitude of the transmitted signal remains constant. Each of the three binary components can be independently controlled in magnitude. In essence, the modulation technique forms the sum of two of the components for modulation onto one phase of the carrier. The third component plus an intermodulation signal forms the quadrature phase component. This is illustrated in the phase plane plot in Figure 17.23. The addition of the intermodulation component to the third signal component as illustrated has to be synchronized to all the other signals in order to form a constant magnitude carrier vector. One feature of Interplex, is that many different intermodulation components are possible. The one that is illustrated in Figure 17.23 is not quite as small as it could be. An optimization process is required to determine the smallest intermodulation component. This is a required step to maximize the efficiency of the multiplex method.

Other possible intermodulation arrangements are possible, for example with components added to both carrier phase components. However, these lack the simplicity and signal decoupling that are illustrated in Figure 17.23. The angles required for the carrier vector have been established in detail for the L1 Galileo transmission, see for example [8] for more details. The minimum intermodulation component has a specific form which can be computed from the product of the three signals.

Grey points: transmitted carrier phases

Transition between phase points not shown, dependent upon transmission filter(s) and may be complex.

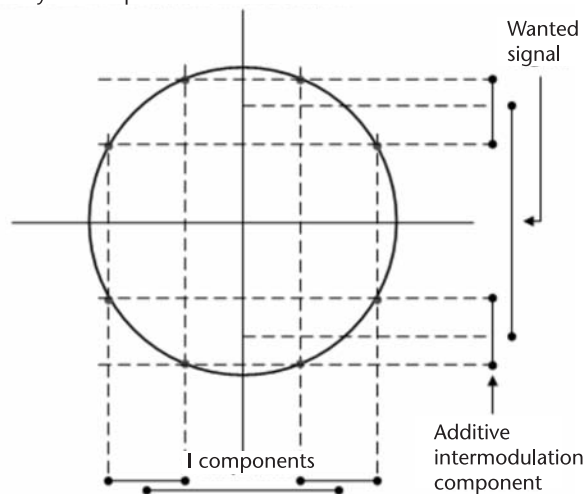


Figure 17.23 Phase plane plot for interplex modulation.

17.4.3 Other Techniques

There are three other known multiplex techniques. These are coherent adaptive subcarrier modulation (CASM); majority voting and intervoting. None have yet been brought into use but one or more certainly will be as the requirement for more signals on the same carrier is now an essential element of the evolution of GPS. The new GPS constellation using GPS III will broadcast four components on L1 with addition of the L1C signal. The current choice for the multiplex method on these satellites is not known or determined.

17.5 Interference

Satellite navigation signals are received at low power levels from transmitters often with output powers of a few tens of Watts, typically in medium Earth orbit (MEO) at distances greater than 20,000 km, and with antenna gains in the 13–14-dB range. Such sources provide signal levels on the Earth's surface of approximately –145 dBW to –160 dBW into a 3-dB linearly polarized antenna. Such signal levels are subject to threats of disturbance from many sources of interference as noted in the Volpe report.

Consequently, the subject of GNSS interference has received considerable attention from international experts during the current evolutionary phase of satellite navigation. In this section, I derive some of the basic interference equations for satellite signal reception and examine the key components.

Satellite navigation signals are provided through international agreement, primarily at the ITU in a number of frequency bands, variously designated as Aeronautical Radio Navigation Services (ARNS) or Radio Navigation Satellite Services (RNSS). In these bands, such signals are offered some protection from interference either as primary or secondary users. At the time of writing, however, ITU agreements have established that any interference shall be below the level that causes “harm.” There are significant disagreements between users and potential users of these bands concerning the definition of harmful interference.

The recent emergence of the International Committee on GNSS (ICG), under United Nations sponsorship, may provide a path to agreement on acceptable interference levels. However, it is already clear that bilateral, let alone multilateral, agreements between sovereign states will be hard to find. The basis for this agreement does not seem to be for the benefit of the ultimate end user.

The ITU does not recognize the difference between military and civil GNSS systems. The ICG is probing what level of international support might exist to respect the needs of sovereign entities to have some protection for their governmental and military satellite navigation signals. While the politics of international agreement on these issues might be hard to predict, the science of interference has at least been well-established. Unfortunately, the criteria and thresholds that constitute “harmful interference” still elude international agreement.

The starting point for the technical analysis is a definition of the wanted satellite transmission, the plethora of noise sources and some structure of the receiver. This system diagram is shown in Figure 17.24 and shows a single wanted satellite source and unwanted signals from the following:

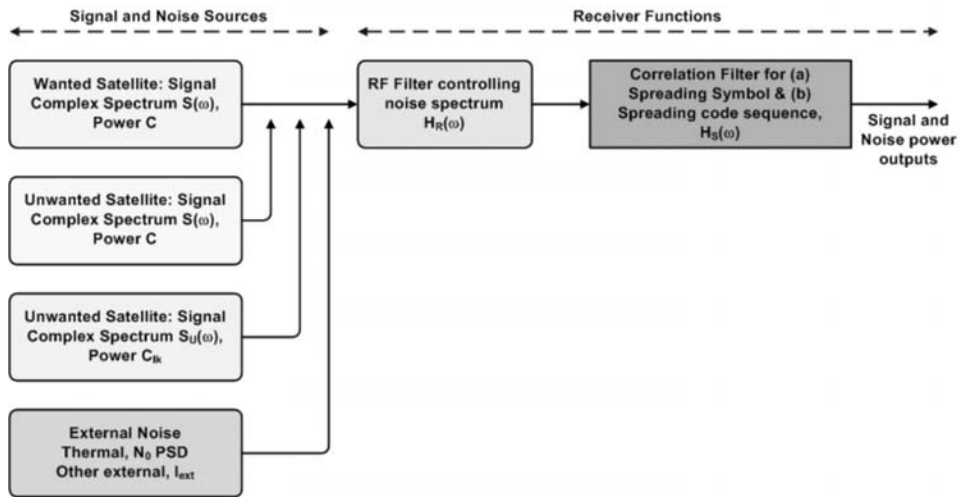


Figure 17.24 Receiver system model for interference computations.

- Other navigation satellites in the same constellation (intrasystem interference) with the same signal structure;
- Other navigation satellites from the same constellation with different signal structure (intrasystem interference);
- Other navigation satellites from different constellation, with similar or different signal structures (intersystem interference), I_{inter} ;
- Background thermal noise, N_0 , often includes the noise from the receiver amplifiers (contained in the description of the noise figure);
- Other noise external sources, I_{ext} .

The receiver is segmented into several processing sections, most represented by filters. There is a receiver filter, $H_R(\omega)$, which removes out-of-band interference or that from other satellites that can be separated by frequency spectrum; a filter for providing a correlation function with the spreading symbol, and finally a correlator for despreading the wanted satellite code sequence. The effects of unwanted signals and noise cause the input equivalent received noise level to be increased. The use of nonideal filters or correlators, such as H_R , often reduce the noise level and almost always reduce the resulting signal output from the correlators. Often, $H_R(\omega)$ can be chosen to selectively mitigate the effects of the (non-white) interference thereby improving the effective carrier to noise ratio.

I note that this is not an “optimum” solution but a pragmatic response to improving performance in the presence of unwanted interference. Often the spectrum of interference is not constant, or the interference may have pulse-type characteristics. In these cases, it may be necessary to seek more complex solutions, such as receiver blanking, adaptive filters (which seek out regions of the spectrum that have excessive noise levels in order to level or excise them), or complex antenna arrangements to steer reception sensitivity away from angularly distinct sources of interference—well-known as controlled reception pattern antennae (CRPAs).

17.5.1 Performance Metrics

Referring to Figure 17.24, the receiver output under the conditions of signal and noise excitation can be computed and summed to find the combined effect since the receiver is a linear network satisfying the principles of superposition. Following this method, the output carrier to noise ratio is found to be:

$$\left. \frac{C}{N} \right|_{\text{eff}} = \quad (17.27)$$

$$\frac{C_S \cdot \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \cdot H_R(\omega) \cdot H_S(\omega) \cdot d\omega \right|_{\text{max}}^2}{\left\{ N_0 \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} |H_S(\omega) \cdot H_R(\omega)|^2 \cdot d\omega + \sum_k C_{Ik} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} |H_S(\omega) \cdot H_R(\omega)|^2 \cdot \Phi_{Ik}(\omega) \cdot d\omega \right\}}$$

The computation has been performed in the frequency domain and requires that the receiver filters are adapted to the transmitted signal spectrum, $S(\omega)$, to maximize signal energy at the output of the receiver. This is the numerator in the equation above. The denominator comprises a (double-sided) thermal noise component of level N_0 (dBW/Hz) whose power at the receiver output is:

$$N_{\text{eff}} = N_0 \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} |H_S(\omega) \cdot H_R(\omega)|^2 \cdot d\omega \quad (17.28)$$

and a second term that is the sum of all other interference sources (shown as k in number). As these are considered as uncorrelated (statistically independent) noise sources, they add on a power basis only:

$$I_{\text{intra}} + I_{\text{inter}} + I_{\text{ext}} = \sum_k C_{Ik} \frac{1}{2\pi} \int_{-\infty}^{\infty} |H_S(\omega) \cdot H_R(\omega)|^2 \cdot \Phi_{Ik}(\omega) \cdot d\omega \quad (17.29)$$

where the summation is over all spectrally defined sources (k) with individual power levels of C_{Ik} and corresponding PSD of Φ_{Ik} . The interference source spectra at the receiver output are modified by the power response of the two receiver filters.

The effective signal-to-noise equation is usually maximized by selecting an optimum frequency, or time, characteristic for the receiver filter, $H_S(\cdot)$, which is then known as the matched filter. In a reference case, the process of optimization assumes that the receiver filter, $H_R(\omega)$, is absent ($H = 1$ everywhere). The optimization uses the Cauchy-Schwarz inequality to provide an upper bound on the integral

in the numerator. The optimization yields the matched filter receiver when the noise background is white (or flat with frequency). With these simplifications the equation becomes:

$$\begin{aligned}
 \left. \frac{C}{N} \right|_{\max} &= \frac{C_S \cdot \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \cdot H_S(\omega) \cdot d\omega \right|_{\max}^2}{\left\{ N_0 \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} |H_S(\omega)|^2 \cdot d\omega \right\}} \\
 &\leq \frac{C_S \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 \cdot d\omega \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} |H_S(\omega)|^2 \cdot d\omega}{\left\{ N_0 \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} |H_S(\omega)|^2 \cdot d\omega \right\}}
 \end{aligned} \tag{17.30}$$

where $H_R(\omega) = 1$, and all $C_{Ik} = 0$.

When the noise background is not white, due perhaps to the addition of an interference component, the Cauchy-Schwarz Inequality still provides an optimum solution but of far more complex form—this is known as a *matched* filter against a colored noise background. I am not aware of any use of this technique in GNSS receivers so far.

The equation above can be further simplified through (a) the cancellation of the integrals of the $H_S(\cdot)$ terms and (b) noting that the integral of $S^2(\cdot)$ is unity by definition. Then, the maximum value of the SNR at the output of the code and spreading symbol correlator is:

$$\left. \frac{C}{N} \right|_{\max} = \frac{C_S \cdot \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \cdot H_S(\omega) \cdot d\omega \right|_{\max}^2}{\left\{ N_0 \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} |H_S(\omega)|^2 \cdot d\omega \right\}} \leq \frac{C_S}{N_0} \tag{17.31}$$

This means that the output of the correlation is, under “matched” processing conditions, an exact equivalent of the carrier-to-noise ratio (CNR) at the receiver input. Consequently, the equation above can be used to provide equivalent *input-referred* signal and noise values due to either imperfect receiver processing [for example to the presence of a nonmatched signal filter, $H_R(\cdot)$] and additional noise sources such as C_I even though their spectra are not white nor the same as the desired signal spectrum.

When the receiver processing is matched to the satellite transmission [apart from $H_R(\cdot)$], another simplification occurs. The matched form for $H_S(\cdot)$ is found from the equality condition specified in the Cauchy-Schwarz inequality; that is:

$$\begin{aligned}
 H_S(\omega) &= \lambda \cdot S^*(\omega) \text{ whereby } \lambda \text{ is scalar, and} \\
 |H_S(\omega)|^2 &= \lambda^2 \cdot |S(\omega)|^2 \\
 &= \lambda^2 \cdot \Phi_S(\omega)
 \end{aligned} \tag{17.32}$$

The input-referred effective CNR can then be rewritten to:

$$\begin{aligned}
 \frac{C}{N} \Big|_{\text{eff}} &= \left\{ \frac{C_S}{N_0} \right\} \\
 &\cdot \frac{\left| \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_S(\omega) \cdot H_R(\omega) \cdot d\omega \right|^2}{\left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_S(\omega) \cdot |H_R(\omega)|^2 \cdot d\omega + \sum_k \frac{C_{lk}}{N_0} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_S(\omega) \cdot |H_R(\omega)|^2 \cdot \Phi_{lk}(\omega) \cdot d\omega \right\}}
 \end{aligned} \tag{17.33}$$

The thermal noise term in the denominator and the signal energy multiplier in the numerator can be recognized as constants dependent upon the energy passed by the receiver filter, $H_R(\cdot)$ and written as:

$$\begin{aligned}
 \eta &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_S(\omega) \cdot |H_R(\omega)|^2 \cdot d\omega \\
 \hat{\eta} &= \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_S(\omega) \cdot H_R(\omega) \cdot d\omega \right|^2
 \end{aligned} \tag{17.34}$$

then

$$\frac{C}{N} \Big|_{\text{eff}} = \left\{ \frac{C_S}{N_0} \right\} \frac{\hat{\eta} \cdot C_S}{\left\{ \eta \cdot N_0 + \sum_k C_{lk} \cdot \kappa_{Sk} \right\}} \tag{17.35}$$

where

$$\kappa_{Sk}^R = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_S(\omega) \cdot |H_R(\omega)|^2 \cdot \Phi_{lk}(\omega) \cdot d\omega$$

The term κ is known as the spectral separation coefficient (SSC). The subscript refers to the sources of the interfering spectrum and the superscript to the identification of the presence of a nonmatched filter, $H_R(\cdot)$ in the receiver chain. The η

terms are called noise and signal (with η) efficiency terms as they refer to the effectiveness of the receiver filter $H_R(\cdot)$ in removing wanted noise or signal energy. It can be shown that the η terms are always less than unity except when the frequency transfer response, $H_R(\cdot)$, passes all signals equally [$H_R(\cdot) = 1$].

17.5.2 Spectral Separation Coefficients (SSC)

The SSC describes the interaction between two noise like signals with different (or the same) spectra:

$$\kappa_{12} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_1(\omega) \cdot \Phi_2(\omega) \cdot d\omega \quad (17.36)$$

and a self-SSC is defined as:

$$\kappa_{11} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_1(\omega) \cdot \Phi_1(\omega) \cdot d\omega \quad (17.37)$$

The spectral separation coefficients can be broadly interpreted as the energy output from a filter of power frequency response, $\Phi_2(\omega)$ (note the unit area requirement), when excited with noise having a spectrum, $\Phi_1(\omega)$ having unit energy.

17.5.2.1 Effects of Band-Limiting Filters

In some cases, a further filter, with frequency transfer function, as before, of $H_R(\omega)$, may be interposed between the noise source and the filter with the $\Phi_2(\omega)$ response. The additional filter modifies the SSC, which is consequently redefined as:

$$\begin{aligned} \hat{\kappa}_{12}^R &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_1(\omega) \cdot |H_R(\omega)|^2 \cdot \Phi_2(\omega) \cdot d\omega \\ \hat{\kappa}_{11}^R &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_1(\omega) \cdot |H_R(\omega)|^2 \cdot \Phi_1(\omega) \cdot d\omega \end{aligned} \quad (17.38)$$

The limits of integration for the SSC integrals are set at $\pm\infty$. This assumes that GNSS transmitters have no filtering. In normal cases, such filtering does exist. The resulting SSC computations require correction for the energy loss in the satellite filters due to their bandwidth limitations. In order to examine some realistic cases, I assume that the infinite bandwidth PSDs (Φ) continue to represent the transmitted power spectrum between the satellite bandwidth limits of $\pm 2\pi\beta_T$ —implying the use of satellite filters with rectangular frequency response. The PSD then requires correction to unit area:

$$\Phi^T(\omega) = \begin{cases} \frac{\Phi(\omega)}{2\pi\beta_T} & \text{for } |\omega| \leq 2\pi\beta_T \\ \frac{1}{2\pi} \int_{-2\pi\beta_T}^{2\pi\beta_T} \Phi(\omega) d\omega & \\ 0 & \text{otherwise} \end{cases} \quad (17.39)$$

When the receiver filter, $H_R(\cdot)$ is modeled by a rectangular filter of one-sided radial bandwidth $2\pi\beta_R$, the SSC definitions can be simplified to:

$$\begin{aligned} \hat{\kappa}_{12}^R &= \frac{1}{2\pi} \int_{-2\pi\beta}^{2\pi\beta} \Phi_1^T(\omega) \cdot \Phi_2^T(\omega) \cdot d\omega \\ \hat{\kappa}_{11}^R &= \frac{1}{2\pi} \int_{-2\pi\beta}^{2\pi\beta} \Phi_1^T(\omega) \cdot \Phi_1^T(\omega) \cdot d\omega \end{aligned} \quad (17.40)$$

where $\beta = \min(\beta_R, \beta_T)$.

These equations contain corrections for the transmitted PSDs (using Φ^T instead of Φ) and for a rectangular receiver filter that restricts the reception bandwidth. The contribution to the integral for frequencies greater than the lowest value of the receiver or transmitter bandwidth is zero.

Tables of κ values are often compiled by those interested in interference computations as these provide ready means to estimate the effects. A few such values are included below for known PSD shapes—see Table 17.1, which has been compiled using known satellite transmission bandwidths of 40.96 MHz for Galileo and 30.69 MHz for GPS. Receiver bandwidths are 24 MHz (double-sided) for most cases and 40.69 MHz for Galileo PRS receivers. The entries in Table 17.1 have been

Table 17.1 SSC Values for Various GNSS Signals and Receivers

Transmitter PSD	Receiver for	<i>Tx Bandwidth</i> (x 1.023MHz)	<i>Rx Bandwidth</i> (MHz)	SSC (dB/Hz)
BPSK-R1 C/A code	BPSK-R1	30	24	−61.8008
BPSK-R10 P(Y) code	BPSK-R10	30	24	−71.2522
BPSK-R10	BPSK-R1	30	24	−69.9071
BPSK-R1	BOC(10,5) M-code	30	24	−87.1116
BOC(10,5)	BPSK-R1	30	24	−87.1116
BOC(1,1)	BPSK-R1	30	24	−67.7619
MBOC(6,1,1/11)	BPSK-R1	30	24	−68.0984
MBOC(6,1,1/11)	BOC(10,5)	30	24	−81.9455
BOC(15,2.5) PRS	MBOC (30.69 MHz)	40	24	−88.8766
BOC(15,2.5)	BPSK-R1	40	24	−108.6910
BOC(15,2.5)	BOC(15,2.5)	40	40.69	−68.4503
BOC(15,2.5)	BPSK-R2 (14 MHz)*	40	40.69	−70.0002

*Note that this is one of three signals planned for emission in the L1 band by the Chinese Compass System and is offset from 1575.42 MHz (band center) by 14 x 1.02 3MHz.

derived from simple computer programs using the known spectral characteristics, bandwidths, and limiting values as identified above.

As an example of interference calculations, an example of C/A code interference with itself is provided. We assume that the thermal noise level is -204 dBW/Hz and that the number of visible satellites emitting C/A code signals is 13, including the wanted satellite. We assume that the average level of received signal is -155 dBW, the aggregate value of received C/A code energy from unwanted satellite signals is $-155 + 10.8 = -144.2$ dBW. The figure of 10.8 dB represents the increase in signal energy due to the 12 (of 13) unwanted satellite signals. The leakage of these signals into the wanted receive channels is then $-144.2 - 61.8008 = -206.01$ dBW/Hz (this is just 2 dB below the thermal noise level).

The aggregate value of the constellation contribution (G_{agg} —aggregate gain) computation requires simulation of the satellite positions over a 24-hour period and the signal levels resulting from them, involving the satellite and receiver antenna gain patterns. The G_{agg} value is not constant over the Earth's surface and this implies further complexity in the detailed analysis. A useful figure for the G_{agg} for GPS is approximately 12.2 dB and is dependent upon satellite constellation numbers. The figure lifts the interference level to -204.6 dBW/Hz causing a near $+3$ dB increase in the background noise floor. This effect will be mitigated by the noise figure of the receiver, but clearly, different receivers with a range of noise figures (NF) will exhibit different susceptibility to this intra-system interference depending upon their NF.

17.5.2.2 Cross-Power Spectral Density

A constituent part of each of the spectral separation coefficients is the product of two power spectral densities:

$$\Phi_{12}(\omega) = \Phi_1(\omega) \cdot \Phi_2(\omega) \quad (17.41)$$

and

$$\hat{\Phi}_{12}^R(\omega) = \Phi_1(\omega) \cdot |H_R(\omega)|^2 \cdot \Phi_2(\omega)$$

We call these functions *cross-power spectral densities*. They have particular utility in designing compatible GNSS signals through establishing the main areas of interaction between two signals (1 and 2), which may be modified by the receiver filter, $H_R(\omega)$. The related spectral separation coefficients are the integrals of the cross-power spectral densities. As an illustration, the cross-power spectral density for the pair of GNSS signals BOC(2,2) and BOC(8,4) is shown in Figure 17.25. Note that the cross PSD has been raised by 60 dB for illustration purposes. From the graph, it can be noted that the main region of interaction is at 8 MHz ($\times 1.023$). Receiver performance for the BOCs(8,4) might be improved by filtering the received signal in the region ± 4 MHz. Since the highest magnitude interaction is in the same part of the band as the BOC(2,2) main lobes, it is unlikely that filtering will significantly improve the performance of the BOCs(2,2) receiver.

17.5.2.3 Time Domain Derivation

An alternative means to derive the SSC can be established using a formulation in the time domain. We have not concentrated upon this approach in the original derivation. However, it can provide a useful means of rapid computation for some examples of satellite PSD. The method often works most easily when there are no bandwidth restrictions.

The satellite PSDs can be equivalently represented by the Fourier transform equivalents in the time domain—these are simply the autocorrelation functions of the spreading symbols. Making this substitution in the SSC equation yields:

$$\begin{aligned}
 \kappa_{12} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_1(\omega) \cdot \Phi_2(\omega) \cdot d\omega \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_1(t) e^{-j\omega t} dt \cdot \int_{-\infty}^{\infty} \phi_2(u) e^{-j\omega u} du \cdot d\omega \\
 &= \int_{-\infty}^{\infty} \phi_1(t) \cdot \phi_2(u) \cdot \delta(t+u) \cdot du dt \\
 &= \int_{-\infty}^{\infty} \phi_1(t) \cdot \phi_2(-t) \cdot dt \\
 &= 2 \int_0^{\infty} \phi_1(t) \cdot \phi_2(t) \cdot dt
 \end{aligned} \tag{17.42}$$

noting that $\phi_2(-t) = \phi_2(t)$ as $\Phi(\omega)$ is a real function.

The integral of the product of two known autocorrelation functions can often be performed by hand; for example, the ACF for a rectangular spreading symbol is triangular on the interval of twice the spreading symbol length. Using this technique for the C/A code spreading symbol, the infinite bandwidth κ_{11} has the value $2\Delta T/3$ ($= -61.8597$ dB/Hz), where ΔT is the spreading symbol duration. Note that the finite bandwidth value (transmitted in 30.69 MHz, received in 24 MHz double-sided) is slightly reduced to -61.8008 dB/Hz.

SSCs with Frequency Offset

There are two reasons why one might require an SSC derived from known spreading code modulations but with an offset in the center frequency between the PSDs. The revised definition takes the form:

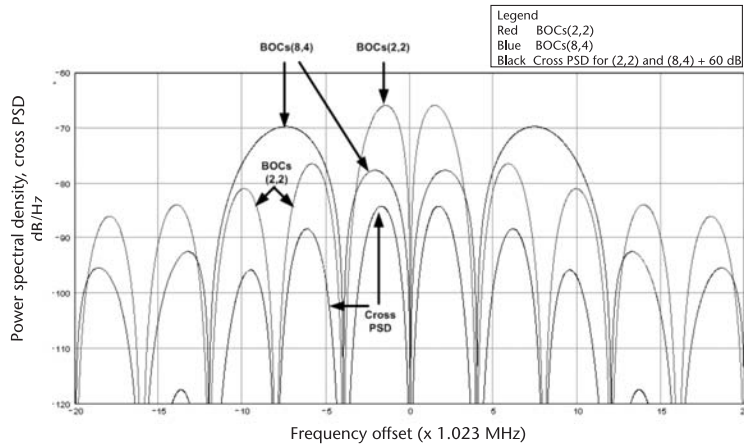


Figure 17.25 BOCs(2,2) and BOCs(8,4) and cross-PSD. (Color image is included on the DVD.)

$$\hat{\kappa}_{12}^R(\omega_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_1(\omega) \cdot |H_R(\omega)|^2 \cdot \Phi_2(\omega - \omega_0) \cdot d\omega \quad (17.43)$$

$$\hat{\kappa}_{11}^R(\omega_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_1(\omega) \cdot |H_R(\omega)|^2 \cdot \Phi_1(\omega - \omega_0) \cdot d\omega$$

Care needs to be exercised in some cases when evaluating the integrals to ensure that the correct frequency limits are used.

The two cases where this modification is required are described as follows. One occurs when the SSC computation involves a Doppler offset in the reception of a specific signal. One clear example is C/A code signals where the detailed SSC has a significant fine frequency structure resulting from the code repetition rate of 1 KHz and a randomizing data component at 50 Hz. This results in a 20-fold repetition of the code sequence for each data bit. We have not examined this special case in this chapter and it is the subject of ongoing research even now. Most of the other GNSS transmissions have set the code repetition interval and each data symbol to have the same duration. This vastly reduces the Doppler sensitivity of the SSC values.

The second case is when one of the components in the SSC calculation has a transmission frequency offset. This can happen when the two SSC components arise from different GNSS systems. One example is the EU Galileo system and the Chinese Compass system. Galileo PRS has a BOC(15,2.5) spreading symbol modulation centered on L1 whereas the Compass system has two QPSK-R2 signals with center frequencies at $\pm 14 \times 1.023$ MHz offset from the band centre. The PSDs are plotted in Figure 17.26 for these signals. The close overlap of spectrum occupancy is clearly evident. Filtering is unlikely to mitigate any mutual interference effects for either Compass or Galileo PRS users.

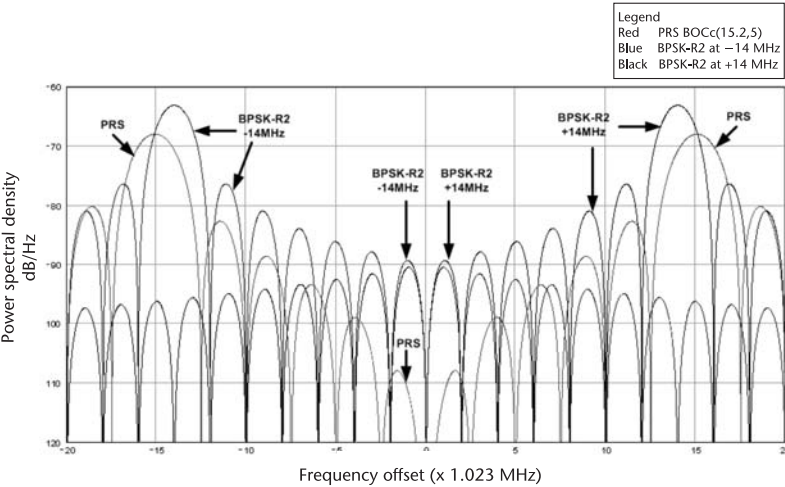


Figure 17.26 Power spectral densities for Galileo PRS and compass BPSK-R2 modulation. (Color image is included on the DVD.)

17.6 Listing of Proposed Systems and Signal Characteristics

The following section contains tables of the characteristics of the present or planned GNSS using CDMA access techniques. This does not include GLONASS although there has been an informal announcement that the Russian Federation is indeed considering the evolution of GLONASS to include a new service with CDMA.

Apart from GPS and GLONASS, the other systems are still using experimental or test satellites and have a clear intention to deploy complete satellite constellations. Regional systems including QZSS, Gagan, and Beidou have not been included. In addition, a further regional system is planned by the Indian government—IRNSS—though this is still in the final system design stages.

There are a number of GPS overlay systems that provide regional coverage, such as WAAS, EGNOS, MSSAS, and MTSAS, whose details are also not included in these tables. WAAS is fully deployed and EGNOS is in the final stages of certification. WAAS is discussed in more detail in Chapter 4.

17.6.1 Global CDMA Satellite Navigation Systems I: GPS

Table 17.2 GPS L1 Signal Characteristics

<i>GPS L1, 1575.42 MHz, CDMA Access Technique, 5° Minimum Elevation</i>					
Service	C/A code	P(Y) code	M Code	L1C	L1C
Component	Data	Data	N/A	Data	Pilot
Spreading Symbol Modulation	BPSK-R1	BPSK-R10	BOCs(10,5)	TMBOC (6,1,1/11)	TMBOC (6,1,1/11)
Subcarrier Frequency (x 1.023 MHz)	—	—	10.0	1.0	1.0, 6.0
Code Frequency (x 1.023 MHz)	1.0	10.0	5.0	1.0	1.0
Code Family	Gold	m-Sequence	N/A	Weil	Weil
Primary PRN Length	1,023	1 week	N/A	10,230	10,230
Secondary PRN Length	—	—	—	—	1800
Data Rate	50 bps	50 bps	N/A	100 bps 50 bps	—
Minimum Received Power (dBW)	−158.5	−161.5	N/A	−157.0	−157.0

Table 17.3 GPS L2 Signal Characteristics

<i>GPS L2, 1227.60 MHz, CDMA Access Technique, 5° Minimum Elevation</i>				
Service	CM Code	CL Code	P(Y) Code	M Code
Component	Data	Pilot	Data	N/A
Spreading Symbol Modulation	BPSK-R1 511.5 KHz time multiplex	BPSK-R1 511.5 KHz time multiplex	BPSK-R10	BOCs(10,5)
Subcarrier Frequency (x 1.023 MHz)	—	—	—	10
Code Frequency (x 1.023 MHz)	0.5	0.5	10.0	5.0
Code Family	m-Sequence	m-Sequence	m-Sequence	N/A
Primary PRN Length	10230 (20 ms)	767250 (1.5s)	1 Week	N/A
Secondary PRN Length	—	—	—	N/A
Data Rate	50 bps 25 bps	—	50 bps	N/A
Minimum Received Power (dBW)	−161.5	−161.5	From −164.5 to −160.0	N/A

Table 17.4 GPS L5 Signal Characteristics

GPS L5, 1176.45 MHz, CDMA Access Technique, 5° Minimum Elevation		
Service	I	Q
Component	Data	Pilot
Spreading Symbol Modulation	QPSK-R10	QPSK-R10
Subcarrier Frequency (x 1.023 MHz)	—	—
Code Frequency (x 1.023 MHz)	10.0	10.0
Code Family	Combined short cycled m-Sequence	Combined short cycled m-Sequence
Primary PRN Length	10230 (1 ms)	10230 (1 ms)
Secondary PRN Length	10	20
Data Rate	100 bps 50 bps	—
Minimum Received Power (dBW)	−157.9	−157.9

17.6.2 Global CDMA Satellite Navigation Systems II: Galileo

Table 17.5 Galileo E1 Signal Characteristics

Galileo E1, 1575.42 MHz, CDMA Access Technique, 10° Minimum Elevation			
Service	OS	OS	PRS
Component	Data	Pilot	Data
Spreading Symbol Modulation	CBOC(6,1,1/11)	CBOC(6,1,1/11)	—
Subcarrier Frequency (x 1.023 MHz)	Two Carriers 1.0 and 6.0	Two Carriers 1.0 and 6.0	—
Code Frequency (x 1.023 MHz)	1.0	1.0	2.5
Code Family	Random Code	Random Code	N/A
Primary PRN Length	4,092	4,092	N/A
Secondary PRN Length	—	25	—
Data Rate	250 bps	—	—
Minimum Received Power (dBW)	−157.0	−157.0	—

Table 17.6 Galileo E6 Signal Characteristics

Galileo E6, 1278.75 MHz, CDMA Access Technique, 10° Minimum Elevation			
Service	CS	CS	PRS
Component	Data	Pilot	Data
Spreading Symbol Modulation	BPSK-R5	BPSK-R5	BOC _c (10,5)
Subcarrier Frequency (x 1.023 MHz)	—	—	10.0
Code Frequency (x 1.023 MHz)	5.0	5.0	5.0
Code Family	Memory Code	Memory Code	N/A
Primary PRN Length	5,115	5,115	N/A
Secondary PRN Length	—	100	—
Data Rate	1,000 sps	—	N/A
Minimum Received Power (dBW)	−155.0	−155.0	—

Table 17.7 Galileo E5 Signal Characteristics

<i>Galileo E5, 1191.795 MHz, CDMA Access Technique, 10° Minimum Elevation</i>				
Service	E5a	E5a	E5b	E5b
Component	Data	Pilot	Data	Pilot
Spreading Symbol Modulation	AltBOC(15,10)	AltBOC(15,10)	AltBOC(15,10)	AltBOC(15,10)
Subcarrier Frequency (x 1.023 MHz)	15.0	15.0	15.0	15.0
Code Frequency (x 1.023 MHz)	10.0	10.0	10.0	10.0
Code Family	m-Sequence	m-Sequence	m-Sequence	m-Sequence
Primary PRN Length	10,230	10,230	10,230	10,230
Secondary PRN Length	20	100	4	100
Data Rate	50 sps	—	250 sps	—
Minimum Received Power (dBW)	−155.0	−155.0	−155.0	−155.0

17.6.3 Global CDMA Satellite Navigation Systems III: COMPASS

Provided by ITU Filing

Table 17.8 Beidou/COMPASS B1 Signal Characteristics

<i>Beidou/COMPASS B1, 1561.098 MHz, CDMA Access Technique</i>			
Service	GSO	N-GSO	GOS and N-GSO
Component	I Data	I Data	Q Data
Spreading Symbol Modulation	QPSK(2)	QPSK(2)	QPSK(2)
Code Frequency (x 1.023 MHz)	2.0	2.0	2.0
Code Family	Concatenated Gold Code	Concatenated Gold Code	Concatenated Gold Code
Data Rate	500 bps	50 bps	500 bps
Minimum Received Power (dBW)	−163.0	−163.0	−163.0

Table 17.9 Beidou/COMPASS B2 Signal Characteristics

<i>Beidou/COMPASS B2, 1207.14 MHz, CDMA Access Technique</i>			
Service	GSO	N-GSO	GOS and N-GSO
Component	I Data	I Data	Q Data
Spreading Symbol Modulation	QPSK(2)	QPSK(2)	QPSK(2)
Code Frequency (x 1.023 MHz)	2.0	2.0	10.0
Data Rate	500 bps	50 bps	500 bps
Minimum Received Power (dBW)	-163.0	-163.0	-163.0

Table 17.10 Beidou/COMPASS B3 Signal Characteristics

<i>Beidou/COMPASS B3, 1268.52 MHz, CDMA Access Technique</i>			
Service	GSO	N-GSO	GOS and N-GSO
Component	I Data	I Data	Q Data
Spreading Symbol Modulation	QPSK(10)	QPSK(10)	QPSK(10)
Code Frequency (x 1.023 MHz)	10.0	10.0	10.0
Data Rate	500 bps	50 bps	500 bps
Minimum Received Power (dBW)	-163.0	-163.0	-163.0

17.7 Summary

This chapter has concentrated on the evolution of GNSS signals from the beginnings in GPS with two BPSK signals on L1 and (usually) one BPSK signal on L2. The evolution has encompassed changes in spreading symbol modulation, which has a major influence on spectral shapes and spectrum occupancy. It is evident that the desire for sovereign powers to own and operate their own GNSS systems exceeds the spectrum available to sustain nonoverlapping signal structures. One might contemplate, perhaps, that the overlapping that exists is purposeful rather than accidental.

We have developed tools to ensure that the reader is equipped to compute the required power spectral densities from the first principles or otherwise as needed. We have seen how signal design has an important effect on multipath sensitivity and the potential for the application of advanced mitigation techniques as well as range measurement accuracy.

We have also covered the subject of interference in some detail providing the most important tools for the computation and engineering of GNSS susceptibility to intrasystem, intersystem, and external interference sources. There is no doubt that this subject will continue to be of major concern for system designers as different GNSS providers compete for space in the ever more crowded GNSS allocated spectrum.

Finally, the chapter has listed the major CDMA-based GNSS systems either deployed or in the process of deployment. The characteristics of Galileo are well-

established with a published signal in space ICD whose parameters are reasonably stable. The data from the Chinese Compass system has been derived from the ITU filing data. At this time, there does not appear to be any stable ICD in the public domain. It is not clear whether the presently known characteristics will change during the current experimental stage.

References

- [1] Betz, J. W., "The Offset Carrier Modulation for GPS Modernization," *Proc. National Technical Meeting, Institute of Navigation ION-NTM 1999*, January 1999, San Diego, CA, pp. 639–648.
- [2] Spilker, J. J., *Global Positioning System, Vol. 1*, eds. B. W. Parkinson, J. J. Spilker, P. Enge, and P. Axelrad, Chapter 3, "GPS Signal Structure and Theoretical Performance," Washington, D.C.: AIAA, 1996.
- [3] Pratt, A. R., "Performance Limits of Multipath Mitigation for Short Delay," *Proc. of the Institute of Navigation, GNSS-ION 2006*, Fort Worth, TX, September 2006.
- [4] Pratt, A. R., and J. I. R. Owen, "Performance of GPS Galileo Receivers Using m-PSK BOC Signals," *Proceedings of the International Technical Meeting of the Institute of Navigation, ION-GNSS 2003*, 9–12 September 2003, Portland, Oregon.
- [5] Pratt A. R., et al, "Tracking Complex Modulation Waveforms—How to Avoid Receiver Bias," *Proceedings of the International Technical Meeting of the Institute of Navigation, IEEE/ION PLANS 2006*, 24–27 April, 2006, Loews Coronado Bay Resort, San Diego, CA.
- [6] Hein, G. W., et. al, "MBOC: The New Optimized Spreading Modulation Recommended for Galileo E1 OS and GPS L1C," *Proc. Int Tech Meeting, IoN, IEEE/IoN PLANS 2006*, 24–27 April 2006, Coronado, San Diego, CA.
- [7] Avila-Rodriguez, J. A., et al, "MBOC: The New Optimized Spreading Modulation recommended for Galileo E1 OS and GPS L1C," ESA Navitec, 11–13 December 2006, Noordwijk, The Netherlands.
- [8] Avila-Rodriguez, J. A., *On Generalized Signal Waveforms for Satellite Navigation*, Ph.D. thesis submitted to Federal Armed Forces University, Neubiberg, Munich (Germany) June 2008.
- [9] Reis, L, "Software Simulation Tools for GNSS2 BOC Signal Analysis," *Proc. International Meeting Io, IoN-GNSS 2002*, September 2002, Oregon Convention Center, Portland, OR.

About the Editors

Scott Gleason received his B.S. in electrical and computer engineering from the State University of New York at Buffalo, an M.S. in engineering from Stanford University, and a Ph.D. from the University of Surrey. He has worked in the areas of GNSS, satellite design, and remote sensing for over 15 years, including at NASA's Goddard Space Flight Center and Stanford's GPS LAAS Laboratory. Over the past decade he has contributed to various GNSS related projects at several companies in the United Kingdom, including SSTL and EADS Astrium. He is currently working for QinetiQ Limited and part time at the National Oceanography Centre Southampton (UK).

Demoz Gebre-Egziabher is an associate professor of aerospace engineering and mechanics at the University of Minnesota, Twin Cities Campus. His research focuses on the design of algorithms and hardware for the navigation and guidance of aircraft, small satellites, and ground vehicles. He is the current secretary of the Satellite Division of the Institute of Navigation (ION) and an associate editor (navigation) for the *IEEE Transactions on Aerospace and Electronic Systems*. He received a B.S. in aerospace engineering from the University of Arizona, an M.S. in mechanical engineering from George Washington University, and a Ph.D. in aeronautics and astronautics from Stanford University.

About the Contributors

Pieter Abbeel received a B.S. and M.S. in electrical engineering from KU Leuven (Belgium) and received his Ph.D. in computer science from Stanford University in 2008. He joined the faculty at UC Berkeley in the Fall of 2008, with an appointment in the Department of Electrical Engineering and Computer Sciences. His research focuses on robotics, machine learning, and control.

Chi O. Ao received his bachelor's degree in physics from the University of California at Berkeley in 1993 and his Ph.D. in physics from the Massachusetts Institute of Technology in 2001. He has worked extensively on the electromagnetic scattering of geophysical random media in the past, and is a coauthor of the book *Scattering of Electromagnetic Waves: Numerical Simulations* (Wiley, 2001). He joined the tracking systems and applications section at the Jet Propulsion Laboratory in 2001 and has been working on various aspects of GPS radio occultation since. He has

made significant contributions in evaluating and improving the quality of the atmospheric retrievals in the lower troposphere. His recent interests include the determination of planetary boundary layer heights from GPS radio occultation and other climate applications.

David M. Bevly is an associate professor in the Department of Mechanical Engineering at Auburn University. He directs the University's GPS and Vehicle Dynamics Laboratory (GAVLAB), which focuses on the control and navigation of vehicles using GPS in conjunction with other sensors, such as Inertial Navigation System (INS) sensors. The GAVLAB's research thrusts are focused towards vehicle dynamics and transportation systems, including heavy trucks, passenger cars, off-road vehicles, as well as autonomous and unmanned vehicles. Dr. Bevly received his B.S. from Texas A&M University in 1995, his M.S. from the Massachusetts Institute of Technology in 1997, and his Ph.D. from Stanford University in 2001, all in mechanical engineering.

Paul Blunt is a senior GPS/GNSS engineer at the GNSS group of Surrey Satellite Technology Limited. He received a master's degree in electrical engineering from the University of Liverpool in 2003 and earned his Ph.D. from the University of Surrey in 2007 for the study of advanced receiver architectures for modernized GPS and Galileo signals. Since then his work has focused on design of transmitter and receiver architectures for Galileo signals and multiantenna beamforming systems for GNSS interference mitigation. His research interests cover all aspects of GNSS receiver design including acquisition, tracking, multipath, and interference mitigation techniques.

Jacob Campbell is an electronics engineer at the Air Force Research Laboratory, Reference Systems Branch, at Wright Patterson AFB. Dr. Campbell received his B.S., M.S., and Ph.D. degrees in electrical engineering from Ohio University where he worked as a research associate for the Ohio University Avionics Engineering Center. While at Ohio University, he interned with Honeywell Labs in the Guidance, Navigation, and Control Center of Excellence, where he performed software GPS research. His research interests include image/shape-aided navigation systems, automated aerial refueling, and other navigation techniques enabling robust navigation in GPS degraded environments.

Richard Fuller, Ph.D., is currently a director of the Wireless Communications Alliance (WCA) and chairman of the Location-Based Services Special Interest Group (LBS SIG), a nonprofit organization dedicated "to promote collaboration, education and knowledge sharing within the Northern California LBS community." In addition to his work at the WCA he cofounded GeoTrax in 2002 and served multiple roles including project manager and technical lead. He has presented numerous speeches and papers and holds several patents in GPS and its applications. In April 2008 GeoTrax was acquired by 3SI Security Systems and Fuller was part of the transition team teaching the 3SI engineers about the Geotrax system. Fuller began his involvement with location-based technologies as chief system architect and programmer at Space Systems/Loral for the GPS Tensor, a position, time, and

attitude space receiver design. Prior to cofounding GeoTrax, Fuller help cofound Traxis where he was vice president of engineering and responsible for directing their products related to the telecommunication and LBS area. Before Traxis, Fuller worked as part of the Wide Area Differential GPS Laboratory at Stanford University, where he studied the use of geostationary satellites as an aircraft landing aid.

Dorota A. Grejner-Brzezinska (Ph.D., 1995, The Ohio State University) is a professor in geodetic science, and leader of the Satellite Positioning and Inertial Navigation (SPIN) Laboratory at The Ohio State University. Her research interests cover GPS/GNSS algorithms, in particular, high precision positioning and navigation. She is vice president of the International Association of Geodesy (IAG) Commission 4, Positioning and Applications, and chair of the Sub-Commission 4.1, Multisensor Systems, and is an IAG fellow; she has been serving on the Institute of Navigation (ION) Council for the past 7 years. She published over 160 peer reviewed journal and proceedings papers, numerous technical reports, and five book chapters on GPS and navigation. She has led over 20 research projects sponsored by DOD, NASA, NGS, NGA, NSF, Federal DOT, and Ohio DOT, with a total budget of over eleven million USD. She is the recipient of the 2005 ION Thomas Thurlow Award, the 2005 United States Geospatial Information Foundation (USGIF) Academic Research Award, and the ESRI Award for Best Scientific Paper in Geographic Information Systems published in 2004. Her work on personal navigation, sponsored by NGA, and was featured as “NGA success story” at the NGA NURI Symposium in Washington DC, in September 2008.

Timo Jokitalo received his M.S. degree in 1994 at the Helsinki University of Technology, Department of Engineering Physics and Mathematics. From 1994 to 2002, he worked at Space Systems Finland Ltd. on various space related embedded and data processing software projects, as well as indoor pseudolite-based navigation. Since 2002 he has worked at Fastrax Ltd., Finland, where his activities have included mainly the research and development of navigation and signal processing algorithms for commercial hardware- and software-based GPS receivers.

Heidi Kuusniemi (born Sandström) received her M.S. degree in 2002 and doctor of technology degree in 2005 from the Department of Information Technology at Tampere University of Technology, Finland. Her doctoral studies on reliability and quality monitoring in personal satellite navigation were partly conducted at the Department of Geomatics Engineering at the University of Calgary, Canada. From 2005 to 2009 she worked in research and development at Fastrax Ltd., Finland, where her field of work included high sensitivity GPS receivers and various reliability enhancement and performance improvement techniques. Since May 2009, Dr. Kuusniemi has been working as a specialist research scientist in the Navigation and Positioning Department at the Finnish Geodetic Institute with research interests including personal positioning and sensor-aided GNSS.

E. Glenn Lightsey holds the W. R. Woolrich Professorship in Engineering in the Department of Aerospace Engineering and Engineering Mechanics at The University

of Texas at Austin. He received his Ph.D. from Stanford University in 1997. Prior to joining The University of Texas in 1999, Dr. Lightsey worked for 13 years as a guidance and control engineer at NASA's Goddard Space Flight Center. Dr. Lightsey specializes in the dynamics and control of spacecraft using avionics sensors such as the Global Positioning System (GPS).

Sherman C. Lo is currently a senior research engineer at the Stanford University Global Positioning System (GPS) Laboratory. He is the associate investigator for the Stanford University efforts on the Department of Transportation's technical evaluation of Loran. He received his Ph.D. in 2002 from Stanford University in developing additional redundancy for GPS in aviation. His current research areas include enhanced Loran, navigation security, and modernized GNSS. He is the recipient of the ION Early Achievement Award (2005) and the International Loran Association (ILA) Presidents Award (2003).

Stephen T. Lowe received B.S. degrees in both physics and mathematics from the Colorado School of Mines, and a Ph.D. in high energy particle physics from Stanford University. He has been employed at the Jet Propulsion Laboratory at the California Institute of Technology since 1987. His research interests include radio interferometry, optical interferometry, experimental tests of general relativity, spacecraft tracking/navigation, GNSS remote sensing, Neutrino astrophysics, Earth-based observations of the Huygens probe descent through Titan's atmosphere, pulsar astronomy, and radar imaging.

Ben Peterson founded Peterson Integrated Geopositioning in 2000. Prior to this he was a U.S. Coast Guard captain and head of the engineering department at the Coast Guard Academy. He is an Academy graduate and earned a Ph.D. in electrical engineering from Yale University. He is a former president of the U.S. Institute of Navigation and an institute fellow. He served as cochair of both the Loran Integrity Performance Panel (LORIPP) and the Loran Accuracy Performance Panel (LOR-IPP), which effectively defined enhanced or eLoran. He currently serves as chair of RTCM SC-127, eLoran Receiver Minimum Performance Standards.

Mark Petovello, Ph.D., is an assistant professor in the Position, Location, and Navigation (PLAN) group in the Department of Geomatics Engineering at the University of Calgary. Since 1998, he has been involved in various navigation research areas including software receiver development, satellite-based navigation, inertial navigation, reliability analysis, and dead-reckoning sensor integration.

Tony Pratt has over 40 year of experience with signal processing and GPS. He graduated with a B.S. and Ph.D. in electrical engineering from Birmingham University, U.K. in 1967. He has held teaching positions at Loughborough University, U.K.; Yale University; IIT, New Delhi; and University of Copenhagen; and holds a full special professorship at the University of Nottingham. His teaching is primarily in signal processing, electronics, probability theory, and satellite navigation system design. He has worked for or consulted Navstar Ltd., Peek, Parthus, QinetiQ Ltd. (U.K.), Cambridge Positioning Systems (now part of the CSR plc group), and

the European Space Agency. Dr. Pratt is a consultant to the U.K. government in the development of the Galileo Satellite System and has played key roles in the signal design and international negotiations. He runs two companies, OrbStar Consultants and OrbStar Ltd., providing various services to the GNSS sector. Dr. Pratt has published numerous papers on signal processing, sonar, and satellite navigation. He published more than 50 papers and holds over 40 patents.

Sam Pullen is a senior research engineer at Stanford University, where he is the director of the Local Area Augmentation System (LAAS) research effort. He has supported the FAA in developing LAAS and WAAS system concepts, requirements, integrity algorithms, and performance models since receiving a Ph.D. degree in aeronautics and astronautics from Stanford in 1996. His current work includes the design of system architectures and algorithms for the next phases of LAAS. He also participates in the development of the next generation of GPS, and his research extends to the broader problem of optimal system design and safety assurance under uncertainty. Dr. Pullen was awarded the Institute of Navigation Early Achievement Award in 1999.

Jason Rife received his B.S. degree in mechanical and aerospace engineering from Cornell University, Ithaca, New York, in 1996, and his M.S. and Ph.D. degrees in mechanical engineering from Stanford University, Stanford, CA, in 1999 and 2004. He is currently an assistant professor of mechanical engineering at Tufts University in Medford, Massachusetts. He directs the Automation Safety and Robotics Laboratory, which applies theory and experiment to characterize the integrity of autonomous vehicle systems. Before joining the faculty at Tufts University, he worked as a research engineer with the Stanford University GPS Laboratory where he specialized in error bounding and fault monitoring. While with the Stanford GPS Lab he directed the Joint Precision Approach and Landing System (JPALS) group, served as a member of the Local Area Augmentation System (LAAS) Integrity Panel, and codesigned the architecture of the Local Airport Monitor (LAM) alternative to LAAS.

Morgan Quigley is a Ph.D. candidate in the Stanford University Computer Science Department. His research endeavors have included software GPS for small aerobatic UAVs, distributed software systems for robotics, and the design and control of low-cost robotic manipulators.

Chris Rizos is currently the head of the School of Surveying & Spatial Information Systems at the University of New South Wales (UNSW). Chris has been researching the technology and applications of GPS since 1985, and established over a decade ago the Satellite Navigation and Positioning group at UNSW, today the largest and best known academic GPS and wireless location technology R&D laboratory in Australia. Chris is the vice president of the International Association of Geodesy (IAG), a member of the Governing Board of the International GNSS Service (IGS), and a member of the IAG's Global Geodetic Observing System Steering Committee. Chris is a fellow of the IAG and a fellow of the Australian Institute of Navigation.

Maarten Uijt de Haag is an associate professor of electrical engineering and computer science and a principal investigator (PI) with the Avionics Engineering Center at Ohio University since 1999. He obtained his M.S. electrical engineering degree from Delft University in The Netherlands in 1994 and his Ph.D. in electrical engineering from Ohio University in Athens, Ohio in 1999. He has authored or coauthored over 80 navigation-related publications, including three book chapters. He is a senior member of the IEEE and AIAA, and a member of the SPIE and ION. Within the latter organization, Dr. Uijt de Haag has served as the air representative on the council and is currently an associate editor for *NAVIGATION: The Journal of the Institute of Navigation*. Furthermore, Dr. Uijt de Haag currently serves on the AIAA Digital Avionics Technical Committee. Dr. Uijt de Haag was awarded the 2008 Institute of Navigation Thomas L. Thurlow Award for his contributions to laser-based navigation and integrity monitors for synthetic vision systems.

Valery U. Zavorotny received his M.S. in radio physics from Gorky State University, Gorky, Russia, in 1971, and his Ph.D. in physics and mathematics from the Institute of Atmospheric Physics of the USSR Academy of Sciences, Moscow, in 1979. He is currently a physicist with the Earth System Research Laboratory of the National Oceanic and Atmospheric Administration (NOAA), Boulder, CO. From 1971 to 1990 he was a research scientist with the Institute of Atmospheric Physics of the USSR Academy of Sciences, Moscow. In 1990 he joined Lebedev Physical Institute, Moscow, Russia. In 1991–2000, he was a CIRES research associate in the Environmental Technology Laboratory of NOAA, Boulder, CO, and became a NOAA/ETL physicist in 2000. His research interests include theory of wave propagation through random media, wave scattering from rough surfaces, and GPS ocean and land remote sensing applications. Dr. Zavorotny is a senior member of the Institute of Electrical and Electronics Engineers, a member of URSI (Commission F), and a member of the American Geophysical Union.

Index

A

- Accuracy, 8
 - in aviation augmentation systems, 252
 - next-generation GNSS, 376
 - range-domain integration and, 281
 - ranging, 10–11
- Acquisition
 - FFT signal, 34
 - GPS software receiver, 133–36
 - output files (debug outputs), 130–31
 - output files (normal output), 130
 - single sideband (SSB), 42
 - weak signals, 294–95
- Additional SF (ASF), 274–75
 - defined, 274
 - generation of, 281–83
 - grid calculation, 281
 - integrity, 286
 - maps, 283
 - spatial and temporal modes, 286
- Advanced forward link trilateration (AFLT), 239, 240
- Aeronautical Radio Navigation Services (ARNS), 468
- A-GNSS, 301
- Agriculture/forestry applications, 7
- Airborne LIDAR Terrain Mapper (ALTM), 203
- Airborne surveys, 369
- Aircraft
 - altimetry measurements, 418–20
 - wind and wave measurements, 421–23
- Aircraft-based augmentation systems (ABAS), 245
 - advantage, 245
 - residual monitor, 262
- Alternative BOC (AltBOC), 461–65
 - carriers, 461–62
 - defined, 461
 - modified, 463
 - modulation, 462
 - PSD for, 464–65
 - signal, 462
 - spectra for Galileo, 465
 - subcarriers, waveshape, 464
- Alternative hypothesis, 314
- Ambiguities
 - CPH, 348, 372
 - fixed, baseline determination with, 358–59
- Ambiguity DOP (ADOP), 187
- Ambiguity resolution, 357
 - algorithms, 353
 - INS and, 187–88
 - Multiple Carrier (MCAR), 376–77
 - overview, 187
 - Triple Carrier (TCAR), 376
- Analog-to-digital converters (ADCs), 28
- Angle of arrival (AOA), 213
 - determination basis, 217
 - location by, 217–19
- Angular displacement, 61
- Angulation, 227–28
 - lateration combined with, 228–29
 - measurement unit vectors, 229
- Antennas
 - controlled reception pattern (CRPAs), 469
 - direction, 332
 - RF system, 240
- Apple Mac OS, 123
- Application-specific integrated circuits (ASICs), 12, 13

Assisted-GPS (AGPS), 144, 301
 Atmospheric profiles, 388–90
 Atmospheric retrievals, 384–90
 derivation of atmospheric profiles, 388–90
 derivation of bending angle profiles, 385–87
 ionospheric calibration, 387–88
 Atmospheric sensing, 381–95
 Attitude
 determination, 338–39
 GNSS/INS integration and, 149
 update equations, 153–55
 Attitude-heading reference systems (AHRS), 149
 Automatic gain control (AGC), 28
 Automotive applications, 7
 Availability, 8, 253
 bending angle profiles, 389
 error bounds and, 254
 next-generation GNSS, 377
 range-domain integration and, 280
 Aviation applications, 7, 245–65
 augmentation system classes, 245–47
 backup navigation capabilities, 251–52
 GNSS modernization, 249–51
 GNSS navigation future, 249–52
 GPS and augmentation benefits, 247–49
 NextGen, 251
 oceanic flight, 247–48
 overland flight, 248
 precision approach and landing, 248–49
 Aviation augmentation systems, 252–64
 accuracy, 252
 availability, 253
 continuity risk, 253
 error bounding (anomalous conditions), 257–61
 error bounding (nominal conditions), 253–57
 integrity risk, 252–53
 monitoring, 261–64
 system performance requirements, 252–53

B

Baseline processing, 356–59
 ambiguity resolution, 357
 with fixed ambiguities, 358–59
 Base station, 347
 Bayesian state estimation, 232–34
 Bayes' theorem, 232
 Bending angle profiles, 385–87
 availability, 389
 illustrated, 388
 Binary coded symbol (BCS), 459
 concept, 459
 LOC versus, 460
 Binary offset carrier (BOC), 445–59
 Alternative (AltBOC), 461–65
 autocorrelation function for, 453
 binary subcarrier modulation for, 446
 composite (CBOC), 455–59
 concepts and definition, 445–47
 discriminator, 40
 early minus late discriminator curve, 42
 multiplex modulation, 453–55
 receivers, 23, 43
 search correlation comparison, 43
 sidebands, 42
 sine-phased, 448, 450
 single sideband (SSB), 42
 spectral properties, 447–51
 spreading symbols, 453
 subcarrier cancellation (SCC), 42
 time multiplex (TMBOC), 457–59
 tracking with BJ algorithm, 46–47
 tracking with DE, 48–53
 tracking with MGD, 44–46
 tracking with SSB, 44
 See also BOC signals
 Binary phase shift keying. *See* BPSK modulation
 Bistatic land and ice sensing, 424–32
 history and applications, 424–25, 427–28
 spacecraft-detected, 426–27, 428–32
 Bistatic radar, 410
 Bistatic radar cross-section (BRCS), 413–14

- Bistatic scattering
 - from land, 416–17
 - from sea ice, 417–18
 - surface, 412
- Bit-error-rate (BER), 213, 221, 455
- Bluetooth, 216, 217
- BOC signals, 39–53
 - correlation peak, 41
 - creation of, 25
 - defined, 39–40
 - modulation, 25–26
 - multiplexed (MBOC), 25–26
 - power spectral density, 26
 - searching for, 39–42
 - tracking, 42–53
 - See also* Binary offset carrier (BOC)
- BPSK modulation, 24, 439, 442–45
 - method and sample waveforms, 442
 - rectangular, 443
- Bump-jump (BJ) algorithm
 - BOC tracking with, 46–47
 - defined, 46
 - false-lock example with, 47
 - peak correction, 47
 - very early (FE) and, 46, 47
 - very late (VL) and, 46–47
- Bumpy discriminators, 46
- C
- C/A code. *See* Clear Acquisition signal
- Carrier phase (CPH), 347
 - ambiguities, 348, 372
 - differenced measurements, 354
 - dual-frequency techniques, 348
 - integer, ambiguity, 353
 - measurements, 347
 - network implementation, 366
 - range measurement, 354
 - smoothing, 348
 - tracking, 365
- Carrier-phase DGNSS (CDGNSS), 92–93
 - defined, 92
 - use of, 92–93
- Carrier phase positioning, 67
- Carrier-to-noise ratio (CNR), 213, 221, 471
- Cauchy-Schwartz inequality, 470–71
- CDMA signals, 239, 240
- Cellular location system, 239–40
- Cellular positioning, 302
- Centroid/center of mass, 232
- Challenging Minisatellite Payload (CHAMP), 2, 382, 391, 392
- Clear Acquisition signal (C/A code), 438
 - Gabor bandwidth for, 452
 - GPS, 27, 33
 - interference, 475
 - PSD for, 445
- Clock
 - anomalies, 103–6
 - bias, estimating, 56–62
 - drift, 64–66
 - errors, 96
 - stability, 295–96
- Closed-loop (CL) tracking, 392
- Code-carrier divergence (CCD), 102–3, 110
- Code division multiple access (CDMA), 250, 441
 - global satellite navigation systems I, 479–80
 - global satellite navigation systems II, 480–81
 - global satellite navigation systems III, 482–83
- Code signal deformation, 98–101
- Coherent adaptive subcarrier modulation (CASM), 468
- Communication, navigation and surveillance, 248
- COMPASS constellation, 2–3, 477, 482–83
- Composite BOC (CBOC), 455–59
- Configuration file, 124–29
 - aiding data, 129
 - data file, 126
 - ephemeris file, 129
 - frequencies, 126
 - logging output, 128–29
 - processing time, 128
 - satellite search strategy, 126–28
 - tracking loop gains, 129
 - WAAS ionosphere corrections, 129
 - See also* GPS software receiver

- Conterminous United States (CONUS), 109–11
- Continuity risk, 8, 253
- Continuously operating reference stations (CORS), 360–67
 - defined, 360
 - IGS infrastructure, 361–64
 - national infrastructure, 364–67
 - next-generation, 365–67, 377–78
 - services, 366–67
 - today, 364–65
- Controlled reception pattern antennas (CRPAs), 469
- Correlation averaging, 407–8
- Correlation power loss, 32
- COSMIC satellites, 384, 391, 394
- Costas PLL discriminators, 34–36
 - carrier phase error, 34
 - characteristic, 35
- Cramer-Rao lower bound (CRLB), 451, 452
- Cross-correlation, 293
 - coping with, 297
 - threshold, 297
- Cross-power spectral density, 475
- Cumulative distribution function (CDF), 255
- Cycle slip, 356
- Cycle slips, 108–9
- D**
- DART, 337
- Deep GNSS/INS integration, 164–65
 - defined, 164
 - illustrated, 164
 - loose integration to, 165
 - See also* GNSS/INS integration
- Delay, 403
 - Doppler maps, 408–10, 423
 - waveforms, 408–10
- Delay-locked loop (DLL), 37–39
 - carrier loop to, 37
 - delay estimate, 50
 - discriminators, 37–39
 - settling time, 52
- DELPHINS tunnel display, 207–8
- Deterministic observability, 184
- DGNSS integrity threats, 97–114
 - faults and anomalies, 97
 - GNSS faults and, 98–108
 - satellite clock anomalies, 103–6
 - satellite code-carrier divergence, 102–3
 - satellite code signal deformation, 98–101
 - satellite ephemeris failures, 106–8
 - satellite low signal power, 101–2
 - signal propagation anomalies, 109–14
 - from system faults, 108–9
- DGPS
 - corrections, 92
 - errors, 92
 - local-area baseline, 90
- Differencing modes, 354–55
- Differential corrections, 143–44
- Differential GNSS (DGNSS), 3, 87–115
 - accuracy improvement, 114
 - carrier-phase, 92–93
 - correction distribution, 94–96
 - correction latency, managing, 96–97
 - defined, 3
 - degree of spatial correlation, 89–93
 - error sources, 89–93
 - fundamentals of, 87–97
 - illustrated, 4
 - integrity improvement, 114
 - introduction to, 87
 - local-area, 88, 93–94
 - reference receivers, 89
 - residual ephemeris error, 90
 - summary, 114–15
 - system faults, 108–9
 - wide-area, 91, 93–94
 - See also* DGNSS integrity threats
- Difficult signal conditions
 - adaptive algorithms, 305–6
 - algorithms for, 303–6
 - constraints on user motion, 304–5
 - defined, 303
 - map matching, 305
 - navigation continuity and, 303
 - See also* Weak signal navigation
- Digital elevation model (DEM), 193
- Digital TV broadcasts, 302

- Dilution of precision (DOP)
 - ambiguity (ADOP), 187
 - geometry (GDOP), 63, 64, 79, 83
 - horizontal (HDOP), 63
 - impact of, 63–64
 - position (PDOP), 63
 - vertical (VDOP), 63, 64
- Direction-finding (DF), 213
 - determination basis, 217
 - location by, 217–19
- Direct Sequence Spread Spectrum (DS-SS), 23
- Distance estimation, 222–23
- Distance measuring equipment (DME), 248
- Doppler, 213
 - advantage, 221
 - calculations, 240
 - effect, 219
 - frequency shift, 403
 - interference-based velocity, 221
 - location estimation, 220
 - location with, 219–21
 - simulated measurements, 71
 - spreading, 403, 424
- Dot-product discriminator, 38
- Double-differencing mode, 354, 355
- Dual estimator (DE)
 - BOC tracking with, 48–53
 - FLL implementation, 49
 - multipath error envelope, 52
 - multipath performance, 51
 - PLL implementation, 49
 - receiver schematic, 49
 - three-loop receiver, 48
- Dual-frequency measurements, 70, 80–82
- E
- eLoran, 271–72
 - defined, 271
 - as global standard, 271–72
 - Loran-C versus, 271
 - for maritime and timing applications, 272
 - See also* Loran
- Ephemeris
 - failures, 106–8
 - file, 129
 - GPS software receiver, 144
 - residual error, 90
- Error bounding
 - anomalous conditions, 257–61
 - excessive acceleration fault, 260–61
 - fault-mode protection levels, 257–59
 - GBAS reference-receiver fault, 259–60
 - nominal conditions, 253–57
 - range-domain, 254
- Error sources, 67–69
- Estimation methods, 226–34
 - Bayesian state, 232–34
 - with nearest neighbor, 229–31
 - nonranging-based location, 231–32
 - probabilistic, with centroid/center of mass, 232
 - with triangulation, 226–29
 - See also* RF systems
- European Centre for Medium-range Weather Forecasts (ECMWF), 390, 391
- Evolution, motivation for, 439–40
- Excessive-acceleration (EA) fault, 260–61
- Executive monitor (EXM), 262
- Exhaustive grid search, 198–99
- Extended Kalman filter (EKF), 67
 - in closed-loop configuration, 166
 - for GNSS/INS, 165–68
 - measurement of observation model, 166
- F
- Fast Fourier Transform (FFT), 33
 - inverse, 136
 - signal acquisition, 34
 - signal detection, 33–34
- Fastgps** software receiver. *See* GPS software receiver
- Fault detection
 - global test, 314
 - local test, 315
- Fault detection and exclusion (FDE), 284
 - example, 322–23
 - iterative procedure, 316
 - in Kalman filtering, 316
 - parameters for, 314–15

- Fault modes and effects analysis (FMEA), 102
- Faults
 - DGNSS system, 108–9
 - excessive acceleration, 260–61
 - GBAS reference-receiver, 259–60
 - in GNSS, 97
 - integrity threats and, 98–108
 - in reference station equipment, 97
 - satellite clock anomalies, 103–6
 - satellite code-carrier divergence, 102–3
 - satellite code signal deformation, 98–101
 - satellite ephemeris failures, 106–8
 - satellite low signal power, 101–2
 - signal propagation, 109–14
- Fault tree, 258
- Field-programmable gate array (FPGA)
 - technology, 13
- Filtering/estimation algorithms, 165–69
- Final operational constellation (FOC), 437
- Fingerprinting
 - defined, 223
 - RF, 213
- FlightGear
 - outputs, 77
 - waypoint navigation with, 77–80
- Formation flying, 336–37
- Forward error correction (FEC), 455
- Forward link, 239
- FPGA-based receivers, 13–14
 - benefits, 14
 - disadvantage, 13
- Frequency domain multiple access (FDMA), 250
- Frequency-locked loop (FLL), 34, 36–37
 - discriminators, 36–37
 - high-bandwidth, 139
 - implementation of DE, 49
 - switch interval, 139
 - tracking, 34
- G
 - Gabor bandwidth, 451–52
 - for C/A, 452
 - MBOC, 454
 - Galileo system, 440
 - AltBOC spectra for, 465
 - constellation, 2
 - global CDMA satellite navigation systems II, 480–81
 - modulation waveforms, 456
 - Public Regulated Service (PRS), 451
 - satellites, adding, 82–83
 - signal availability, 10
 - TLE file, 83
 - Gaussian distribution, 414
 - General positioning system case study, 202–5
 - Geodesy, 349–50, 372–75
 - accuracy requirement, 350–51
 - application, 372–73
 - defined, 349–50
 - methodology, 373–75
 - mission, 350
 - processing model, 354
 - satellite techniques, 350
 - scenario, 374
 - space, 361
 - Geodesy/surveying applications, 7, 347–79
 - baseline processing, 356–59
 - differenced observations, 354–56
 - mathematical models, 351–56
 - network processing, 359–60
 - in next-generation GNSS, 376–79
 - one-way observations, 351–54
 - operational modes and, 367–75
 - technical overview, 350–60
 - Geometric observability, 184
 - Geometry DOP (GDOP), 63, 64
 - change over time, 64
 - recalculated, 79, 83
 - Geometry matrix, 309
 - Global Earth Observing System of Systems (GEOSS), 364
 - Global Geodetic Observing System* (GGOS), 364, 379
 - Global navigation satellite systems. *See* GNSS
 - Global positioning system. *See* GPS
 - Global test, 311–12
 - chi-square distributions related to, 311

- defined, 311
- failure, 314
- fault detection, 314
- GLONASS, 2, 250
 - evolution, 478
 - operation, 250
- GNSS
 - ambiguity-resolution overview, 187
 - applications, 6–8
 - bistatic radar, 410
 - constellations, 2–3
 - differential (DGNSS), 3, 87–115
 - faults within, 97
 - geodesy, 349–50
 - high-altitude, 339
 - history of, 437–39
 - indoor navigation, 6
 - INS integration, 149–89
 - integrated navigation and, 5–6
 - location-based services, 6
 - modernization, 249–51, 340–41
 - network-assisted (A-GNSS) navigation, 3–5
 - next-generation, 376–79
 - nonpositioning uses, 367
 - positioning performance measures, 8–9
 - precise navigation with, 1
 - primary application, 56
 - receivers. *See* Receivers
 - relative, 88–89
 - RTK positioning, 92
 - satellite hardware, 252
 - signals. *See* Signals
 - stand-alone satellite navigation, 3
 - surveying, 348–49
 - updates, 170
 - user architectures, 3–6
 - weak signal supportive systems, 301–3
- GNSS Evolutionary Architecture Study (GEAS), 250
- GNSS/INS integration, 149–89
 - architecture overview, 160
 - attitude solution, 154
 - block labeled algorithm, 165
 - case studies, 177–89
 - concepts, 159–65
 - deep, 164–65
 - error statistics, 159
 - extended Kalman filter (EKF) for, 165–68
 - filtering/estimation algorithms, 165–69
 - implementation, 169–72
 - INS to aid high-accuracy GNSSs, 186–88
 - lever arm and, 173
 - loose, 160–62
 - low-cost GNSS/INS navigator, 177–80
 - MATLAB-based, 178
 - motivation, 159–60
 - practical considerations, 172–74
 - software examples, 188–89
 - step-by-step, 172
 - summary, 174–75
 - tight, 162–63
 - time evolution of, 168–69
 - timing requirements, 173–74
 - vehicle sideslip estimation case study, 180–86
 - See also* Inertial navigation system (INS)
- GNSS integrity channel (GIC), 250
- GNSS-only RAIM, 246
- GNSS simulator, 69–84
 - defined, 69
 - displays from, 76
 - dual frequency calculation, 80–82
 - examples, 74–84
 - Galileo satellites and, 82–83
 - in Google Earth, 79
 - input files, 71–72
 - interface files, 71–73
 - measurement details, 69–71
 - output files, 72–73
 - postprocessing output files, 73–74
 - pseudorange measurements, 69–71
 - simple navigation example, 74–75
 - spacecraft-based receiver, 84
 - summary, 84–86
 - traveling between destinations, 75–77
 - waypoint navigation with FlightGear, 77–80
- GNU/Linux, 122
- Google Earth, 79

- GPS, 1
 - additional frequencies, 10
 - assisted (AGPS), 144
 - bistatic radar, 338
 - C/A code, 27, 33
 - global CDMA satellite navigation systems I, 479–80
 - history of, 437–38
 - kinematic (KGPS), 205
 - OCS, 103
 - overlay systems, 478
 - processing services, 375
 - pseudorange measurements, 57
 - range measurements, 308
 - relative navigation and, 238
 - Standard Positioning Service (SPS), 89
 - time, 66, 132–33
 - See also* GPS software receiver
- GPS and GEO augmented navigation system (GAGAN), 246
- GPS/Meteorology (GPS/MET), 382
- GPS software receiver, 121–46
 - acquisition, 133–36
 - acquisition algorithm, 135
 - acquisition log file, 130–31
 - acquisition steps, 135
 - additional signal processing, 145
 - architecture, 131–45
 - background, 121
 - carrier-phase precise positioning, 145–46
 - carrier-smoothing code measurements, 145
 - command characters, 126–29
 - configuration file, 124–29
 - core modules, 131
 - defined, 121
 - development environments, 122–23
 - differential corrections, 143–44
 - ephemeris data-decoding functions, 141
 - example data sets, 123–24
 - GUI improvement, 145
 - inverse FFT result, 136
 - Kalman filter positioning, 146
 - license, 122
 - main processing loop, 133
 - navigation, 142–45
 - navigation data decoding, 139–41
 - navigation solution processing, 142
 - output display, 123
 - output file, 130–31
 - pseudorange measurements, 134, 143
 - PVT estimation, 145
 - satellite information from aiding data, 144–45
 - satellite information from ephemeris, 144
 - satellite pseudorange measurement, 134
 - software correlator, 137–38
 - speed improvements, 145
 - timing and clock management, 132–33
 - tracking, 136–41
 - tracking loops, 138–39
 - using, 124–31
 - velocity estimates, 126
 - See also* GPS
- GRACE, 337
 - high-level system processing, 342, 343
 - NASA/JPL Blackjack receiver, 341
 - orbits, 394
 - processing raw measurements from, 341–44
 - ultra stable oscillators (USOs), 384
- Gradient-based search, 200–202
 - algorithm, 200–201
 - equation modification, 200
 - for minimum error, 201
 - use of, 201
- Ground-Based Augmentation System (GBAS), 95
 - bandwidth availability, 257
 - broadcast structure of, 247
 - corrections, 96
 - excessive acceleration monitor, 262–64
 - message structure, 95
 - reference-receiver fault, 259–60
 - VHF broadcasts, 96
- Ground-Based Regional Augmentation System (GRAS), 95–96, 245
 - as hybrid, 247
 - initial operating capability, 247
- Group repetition interval (GPI), 272

H

Hamming distance, 232
 Hazardously misleading information (HMI), 245, 246
 Heads-down-display (HDD), 208
 High-altitude elliptical orbit (HEO), 331
 High-altitude GNSS, 339
 High-power amplifier (HPA), 459
 Horizontal DOP (HDOP), 63
 Hypothesis testing table, 314

I

Ice reflections, 427–32
 history and applications of, 427–28
 signal response, 431
 spacecraft-detected, 428–32
 summary, 432
 Indoor navigation, 6
 Industrial, scientific, and medical (ISM), 237
 Inertial measurement unit (IMU), 150
 bias error model parameters, 171
 data processing, 172
 low-cost, integrated with GNSS, 170
 rotation rates, 173
 sensor error models, 169–72
 velocity measurements, 196
 velocity translation, 173
 Inertial navigation system (INS), 149–89
 ambiguity resolution and, 187–88
 body frame, 151
 coordinate frames, 151
 defined, 150
 Earth frame, 151
 error model, 157–58
 for high-accuracy GNSSs, 186–88
 inertial frame, 151
 mechanization equations, 152–57
 navigation frame, 151
 system initialization, 157
 updating, 173
 See also GNSS/INS integration
 Initial operational capability (IOC), 437
 Input-referred effective CNR, 472
 Input-referred signals, 471
 Instrument Landing System (ILS), 95
 Integer CPH ambiguity, 353

Integrated navigation, 5–6
 Integrity monitoring, 306–18
 alternative hypothesis, 314
 FDE in Kalman filtering, 316
 FDE parameters, 314–15
 global test, 311–12
 introduction to, 306–7
 local test, 312–14
 multiple outliers, 315–16
 null hypothesis, 314
 RAIM and, 306–7
 reliability testing, 307–8
 residuals and redundancy, 310–11
 weighted least-squares notation, 308–10
 Integrity monitors, 261
 Integrity risk, 8, 252–53
 Integrity threats. *See* DGNSS integrity threats
 Interference, 468–78
 C/A code, 475
 performance metrics, 470–73
 receiver system model for, 469
 SSC, 473–78
 International DORIS Service (IDS), 350
 International Earth Orientation Service (IERS), 441
 International GNSS Service (IGS), 361–64
 defined, 361
 in future, 363–64
 global distribution of stations, 363
 operation, 361
 product summary, 362
 International Space Station (ISS), 333
 International terrestrial reference frame (ITRF), 155, 350, 441
 Interplex, 467
 Inverted range, 236
 Ionosphere errors
 impact of, 62–63
 size of, 62
 Ionosphere grid points (IGP)
 delays, 143, 144
 geographic box, 143
 Ionosphere pierce point (IPP), 110
 Ionosphere term, 68

- Ionospheric calibration, 387–88
- Ionospheric delay gradients, 109–10
- Ionospheric spatial-decorrelation
 - anomalies, 109–13
 - near-worst-case, 110
 - risk mitigation options, 113
 - threat model parameter bounds, 111
- J**
- Joint tactical information distribution system (JTIDS), 238
- K**
- Kalman filters
 - divergence of, 318
 - extended (EKF), 67, 165–68
 - FDE and, 316
 - inertial, 194
 - integration, 235
 - unscented (UKF), 165
- Kalman gain matrix, 167
- Kinematic GNSS techniques, 367, 372
- Kinematic GPS (KGPS), 205
- Kirchoff approximation and geometric optics limit (KA-GO), 401
- KissFFT routine, 135
- K-mean clustering, 232
- K-nearest-neighbor estimation plots, 241
- L**
- LADAR-based navigation, 192–210
 - exhaustive grid search, 198–99
 - inertial velocity error estimation, 202
 - position estimate and SSE surface, 196–98
 - TERRAIN integration methodology, 192–95
 - terrain-referenced position estimation, 196–202
- Land reflections, 424–27
 - history and applications, 424–25
 - spacecraft-detected, 426–27
 - summary, 432
- Landsat, 400
- Land surveys, 368
- Laser radar (LADAR), 191
 - aiding, 195
 - certifiable positioning system, 194
 - horizontal position estimate error, 205
- Laser scanner optics (LSO) reference point, 196
- Lateration, 226, 228–29
- L-band GNSS frequency allocations, 25
- Least-squares ambiguity decorrelation adjustment (LAMBDA), 357
- Least-squares integration, 234–35
- Lever arm, 173
- Light detection and ranging (LIDAR), 191
 - accuracy requirements, 193
 - Airborne Terrain Mapper (ALTM), 203
 - NASA Dryden DC-8 cargo bay installation, 204
- Linear offset carrier (LOC), 459
 - BOC spectra versus, 460
 - disadvantages, 459
 - implementation, 459
- Lines of positions (LOPs), 270, 271
- Link quality indicator (LQI), 221
- Local Area Augmentation System (LAAS), 92
 - B-value calculation for, 108–9
 - category-I landing certification, 249
 - FAA, 249
 - GAST-C version, 113
 - ground facility (LGF), 110
 - ionosphere spatial anomaly threat model, 110
 - minimum pseudorange correction, 107
 - provable safe prototype (PSP), 112
 - test prototype (LTP), 103
 - tropospheric decorrelation, 114
 - vertical position errors, 112
- Local-area DGNSS
 - defined, 88, 93
 - geometry, 88
 - wide-area DGNSS versus, 93–94
 - See also* Differential GNSS (DGNSS)
- Localizer performance with vertical guidance (LPV), 248
- Local test, 312–14
 - defined, 311
 - erroneous measurement detection, 313

- fault detection, 315
 - normal distributions related to, 313
- LocataNet, 238
- Location-based services, 6
- Location system alternatives, 211–13
- Loose GNSS/INS integration, 160–62
 - defined, 160
 - feedback path, 161
 - illustrated, 161
 - key feature, 161–62
 - open-or-closed-loop configurations, 160–61
- See also* GNSS/INS integration
- Loran
 - additional SF (ASF), 274–75
 - defined, 269
 - eLoran, 271–72
 - group repetition interval (GPI), 272
 - integrity, improved accuracy for, 286
 - lines of positions (LOPs), 270, 271
 - Loran-C, 269–70
 - nominal emission delay (NED), 272, 273
 - phase code interval (PCI), 272–73
 - primary factor (PF), 274
 - secondary factor (SF), 274
 - signal transmission, 272
 - system area monitor (SAM), 274
 - theory of operation, 272–75
 - time of transmission (TOT), 274
 - worldwide coverage, 270
- LORAN/GNSS integration, 269–88
 - historical reasons for, 275
 - position-domain, 276–78
 - range-domain, 278–86
 - scenarios, 276–87
 - tracking loop domain, 287
- Low-cost GNSS/INS navigator, 177–80
 - attitude history, 179
 - ECEF coordinate estimate, 182
 - GNSS outage effect, 181
 - ground trajectory, 178
 - sensor bias estimate, 180
- See also* GNSS/INS integration
- Low-noise amplifiers (LNAs), 28
- Low signal power, 101–2

M

- Mahalanobis distance, 231
- Map matching, 305
- Marine applications, 7
- Marine surveys, 368–69
- Matched filters, 471
- MATLAB/Octave script, 73, 75, 81, 123, 322
- Maximum error range requirement (MERR), 101
- Maximum ionosphere error in vertical position (MIEV), 112, 113
- Mechanization equations, 152–57
 - attitude update, 153–55
 - defined, 152
 - position update, 156–57
 - velocity update, 155–56
- See also* Inertial navigation system (INS)
- MEMS sensors, 302
- Microsoft Windows, 123
- Minimum detectable errors (MDEs), 107
- Monitors
 - ABAS residual, 262
 - executive (EXM), 262
 - GBAS excessive acceleration, 262–64
 - integrity, 261
 - performance, quantifying, 264
 - types of, 262
- See also* Aviation augmentation systems
- Most likely subset (MLS) threat model
 - parameters, 100
- Multimode receivers (MMRs), 248
- Multipath, 293
 - mitigation, 298–300
 - SNR oscillations and, 299
 - space applications and, 333
- Multiple Carrier Ambiguity Resolution (MCAR), 376–77
- Multiple-correlator signal deformation
 - monitoring, 101
- Multiple dwells, 34
- Multiple-gate discriminators (MGD)
 - BOC tracking with, 44–46
 - bumpy, 46

Multiple-gate discriminators (MGD)
 (continued)
 composite, 46
 defined, 44
 smooth, 46

Multiple hypothesis tracking (MHT), 233

Multiple reference consistency check
 (MRCC), 108–9

Multiplex BOC modulation (MBOC),
 453–55
 defined, 453–54
 Gabor bandwidth, 454
 PSD, 454, 456–57
 structure, 25–26
 See also Binary offset carrier (BOC)

N

National Center for Environmental
 Prediction (NCEP), 390, 391

National Spatial Reference System
 (NSRS), 370

Navigation data decoding, 139–41

Navy Navigation Satellite System
 (NNSS), 437

Nearest neighbor (NN)
 defined, 229
 deterministic estimation with, 229–31
 estimation plots, 241
 ratio, 232
 variation, 230–31

Network-assisted GNSS (A-GNSS)
 navigation, 3–5
 defined, 3
 functioning of, 4
 illustrated, 5

Neural network models, 231

Next-generation GNSS, 376–79
 accuracy, 376
 applications, 378–79
 availability, 377
 continuity, 376
 efficiency of positioning, 376–77
 infrastructure improvements, 377–78
 reliability, 377

NextGen program, 251

Nominal emission delay (NED), 272, 273

Nonranging-based location estimation,
 231–32

Normal gravity vector, 155

North-east-down (NED) navigation
 frame, 152

Null hypothesis, 314

Numerical weather prediction (NWP)
 models, 390, 391

Nyquist filtering, 42

O

Occultation measurements, 382–84

Ocean altimetry, 418–21
 aircraft measurements, 418–20
 motivation, 418
 from space, 420–21

On-line Positioning User Service (OPUS),
 370

Orbit
 geometry, 330–32
 GRACE, 394
 precise determination, 335

Organization, this book, 14–20

Outliers
 multiple, 315–16
 single, assumption of, 315

Output files, 130–31
 acquisition (debug outputs), 130–31
 acquisition (normal output), 130
 navigation (Google Earth output), 131
 navigation (normal output), 131
 tracking outputs, 131
 See also GPS software receiver

Oven-controlled crystal oscillators
 (OCXOs), 296

P

Particle filter (PF), 165

Pattern recognition, 223

Performance metrics, 470–73

Personal navigation applications, 6

Phase code interval (PCI), 272–73

Phase-locked loop (PLL), 34–36
 discriminators, 27, 34–36
 implementation of DE, 49
 updating, 35

Phase-of-arrival (POA), 224

- Phase shift keying (PSK)
 - binary (BPSK) modulation, 24, 439, 442–45
 - correlator structure, 29
 - receivers, 23, 29
 - search correlation comparison, 43
 - See also* PSK signals
 - Position, navigation, and timing (PNT), 271
 - Position, velocity, and time (PVT), 12
 - carrier phase positioning and, 67
 - DOP, 63–64
 - error sources, 67–69
 - estimation, 56–69
 - estimation with EKF, 67
 - GPS software receiver, 145
 - initial conditions for, 59
 - ionosphere errors and, 62–63
 - receiver position and clock bias, 56–62
 - receiver velocity and clock drift, 64–66
 - time estimation, 66–67
 - Position-domain Loran/GNSS integration, 276–78
 - defined, 276
 - drawbacks/benefits, 276–77
 - illustrated, 276
 - integrity benefit, 277–78
 - Position DOP (PDOP), 63
 - Position update equation, 156–57
 - Precise Orbit Determination (POD), 335
 - Precision approach guidance system case study, 205–10
 - PRN codes, 26–28, 319
 - faulty, 323
 - orthogonal, 29
 - sequence, 24, 34
 - tiered, 27
 - Probability density function (PDF), 232, 413
 - Programmable hardware, 332
 - Proximity, 213
 - location by, 214–17
 - operations, 336–37
 - Pseudolites, 235–37
 - defined, 235
 - ground-based, 236
 - infrastructures, 301
 - Pseudorandom noise (PRN)
 - sequence autocorrelation, 27
 - spreading signal, 23
 - See also* PRN codes
 - Pseudorange corrections (PRCs), 95
 - Pseudorange measurements, 57, 58
 - defined, 351
 - dual-frequency, 352
 - GPS software receiver, 134, 143
 - in ionosphere errors, 62
 - noisy, combining, 63
 - random noise on, 70
 - simulated, 69–71
 - use of, 352
 - PSK signals, 28–39
 - searching for, 28–34
 - tracking, 34–39
 - tracking techniques, 23
 - Python script, 73, 75, 81, 123
- Q**
- Quadrature phase shift keyed signal (QPSK), 439, 440, 466
 - Quality control, 316–18
 - geometrical considerations, 316
 - navigation solution, 318
 - practical side, 317–18
 - procedure, 317
 - See also* Weak signal navigation
- R**
- Radio Navigation Satellite Services (RNSS), 468
 - Radio occultation (RO), 381–95
 - atmospheric retrievals, 384–90
 - DVD scripts, 394
 - measurements, 382–84
 - recent advances, 392–94
 - self-calibrating measurements, 382
 - weather and climate applications, 390–92
 - Radio Technical Commission for Maritime Services (RTCM)
 - SC-104, 95
 - Range-domain Loran/GNSS integration, 278–86
 - accuracy and, 281

- Range-domain Loran/GNSS integration
 - (continued)
 - availability and, 280
 - case study, 281–83
 - defined, 278
 - illustrated, 278
 - integrity with, 283–86
 - using, 279–80
 - See also* Loran/GNSS integration
- Range errors, 68–69
- Range-rate corrections (RRCs), 95
- Range term, 68
- Ranging
 - codes, longer, 11
 - higher accuracy, 10–11
 - light detection and (LIDAR), 191, 193, 203, 204
- Ratio-nearest neighbor, 232
- Real-time kinematic (RTK)
 - GNSS positioning, 92
 - mode, 348, 349
- Real-time navigation, 335–36
- Received signal strength (RSS), 222
- Receiver Autonomous Integrity Monitoring (RAIM), 283
 - algorithms, 284, 285
 - basis, 306
 - GNSS-only, 246
 - integrity monitoring and, 306–7
 - Loran measurements for, 285
 - technique, 283
 - weak signals and, 306
- Receivers
 - advanced technology, 12–14
 - block diagram, 12, 28
 - BOC, 23, 43
 - conventional, 12–13
 - data rates, 159
 - FPGA-based, 13–14
 - in launch range design, 340
 - position, estimating, 56–62
 - PSK, 23, 29
 - recalculated velocities of, 78
 - reference, 108–9, 347
 - signal level effect on, 294
 - software-defined, 14
 - spacecraft-based, 84
 - in surveying, 347
 - velocity, estimating, 64–66
- Redundancy matrix, 310
- Reference receiver errors, 108–9
- Reflection(s)
 - geometry, 402–3
 - ice, 427–32
 - land, 424–27
 - point location, 403
- Refractivity, 389
- Relative GNSS, 88–89
- Relative navigation, 238
- Reliability
 - next-generation GNSS, 377
 - testing, 307–8, 314
- Remote sensing, 337–38, 399–434
 - with bistatic GNSS reflections, 399–434
 - bistatic land and ice, 424–32
 - correlation averaging, 407–8
 - delay Doppler maps, 408–10
 - delay waveforms, 408–10
 - DVD data, 432–33
 - introduction to, 399–401
 - ocean altimetry, 418–21
 - with reflected GNSS signals, 401
 - reflection geometry, 402–3
 - signal processing, 403–10
 - specular point calculation scripts, 432–33
 - surface mapping, 405–7
 - surface scattering model, 433
 - theory, 410–18
 - traditional, 400–401
- Response rate (RR), 221
- RF fingerprinting, 213
- RF identification (RFID), 213
 - configuration, 215
 - system, 215
 - tags, 215
 - technology, 302
- RF location
 - AOA, 217–19
 - with Doppler frequency, 219–21
 - estimation with signal strength, 221–23
 - by proximity, 214–17

- by radio DF, 217–19
- by TDOA hyperbolic traces, 225
- with time, phase, and differential TOA, 223–26
- TV-based, 238–39
- RF propagation, 212
- RF systems, 211–41
 - antennas, 240
 - characterization in frequency bands, 213
 - contextual processing, 235
 - estimation methods, 226–34
 - example systems, 235–40
 - integration methods, 234–35
 - Kalman filter integration, 235
 - least-squares integration, 234–35
 - location types and classifications, 213–26
- Round-trip-delay (RTD) measurement, 214
- Runway visual range (RVR), 208
- S**
- SAC-C, 391, 392, 393
- Satellites
 - clock anomalies, 103–6
 - clock term, 68
 - code-carrier divergence, 102–3
 - code signal deformation, 98–101
 - ephemeris failures, 106–8
 - low signal power, 101–2
- Satellite-user geometry, 63–64
- Scientific applications, 7–8
- Sea ice
 - bistatic scattering from, 417–18
 - reflections, 427–32
- Sea surface modeling, 414–16
- Self-synchronizing networks, 237–38
- Sideslip estimation, 181–86
 - categories, 181
 - defined, 180–81
 - with filter-blending, 183
 - hypothetical vehicle trajectories, 284
 - lateral force estimates, 183
 - motivation, 181–83
 - observability, 184–86
 - sideslip angle, 181, 182
- Signal multiplex techniques, 465–68
 - CASM, 468
 - Interplex, 467
 - QPSK, 466
- Signal processing, 403–10
 - averaging consecutive correlations, 407–8
 - delay waveforms/Doppler maps, 408–10
 - detection and surface mapping, 405–7
 - See also* Remote sensing
- Signal propagation anomalies, 97, 109–14
 - ionospheric spatial-decorrelation, 109–13
 - tropospheric spatial-decorrelation, 114
- Signals
 - background, 23–28
 - BOC, 39–53
 - BPSK modulated, 24
 - carrier, modulation, 439
 - CDMA, 239, 240
 - deformation, 98–101
 - FFT detection, 33–34
 - improvements, 9–12
 - PRN, 23
 - PSK, 23, 28–39
 - serial search, 31
- Signal strength, 213
 - distance estimation, 222–23
 - location estimation, 221–23
 - pattern recognition, 223
 - space applications and, 333–34
- Signal-to-noise ratio (SNR), 221, 470
- Single-differencing mode, 354
- Single sideband (SSB)
 - acquisition, 42
 - BOC tracking with, 44, 45
- Smooth discriminators, 46
- Software-defined GNSS receivers, 14
- Space, ocean altimetry from, 420–21
- Space applications, 7, 329–44
 - antenna direction and, 332
 - attitude determination, 338–39
 - environment and, 334–35
 - formation flying and proximity operations, 336–37

- Space applications (continued)
 - GNSS modernization and, 340–41
 - high-latitude GNSS, 339
 - introduction to, 329
 - launch, entry, and landing, 340
 - multipath and, 333
 - operational considerations, 329–35
 - orbit geometry and, 330–32
 - precise orbit determination, 335
 - real-time navigation, 335–36
 - remote sensing, 337–38
 - signal strength and, 333–34
 - size/power and, 332
 - spacecraft velocity and, 330
 - summary, 344
- Space-based augmentation systems (SBASs), 94, 96
 - correction implementation, 94
 - monitors for HMI, 246
 - multifunctional-transport (MSAS), 246
- Spacecraft-based receivers, 84
- Space geodesy, 361
- Spectral separation coefficient (SSC), 472, 473–78
 - band-limiting filters and, 473–75
 - cross-power spectral density, 475
 - defined, 472, 473
 - with frequency offset, 476–78
 - self, 473
 - time domain derivation, 476–78
 - values, 474
 - See also* Interference
- SPOT, 400
- Square-wave cosine function, 35
- Stand-alone satellite navigation, 3
- Static GNSS positioning techniques, 367
- Stop-and-go technique, 371
- Subcarrier cancellation (SCC), 42
- Subcarrier locked loop (SLL), 48
- Sum-of-squared-error (SSE), 196
- Support vector mapping (SVM), 232
- Surface mapping, 405–7
- Surface scattering model, 433
- Surveying, 348–49, 368–72
 - airborne surveys, 369
 - applications, 368–69
 - defined, 347
 - infrastructure, 365
 - kinematic technique, 372
 - land surveys, 368
 - marine surveys, 368–69
 - rapid-static technique, 369–71
 - stop-and-go technique, 371
 - techniques, 369
 - See also* Geodesy/surveying applications
- Synchrolites, 237
- Synthetic aperture radar (SAR), 400
- T
- Tactical air navigation (TACAN), 248
- Taylor series expansion, 59, 227
- Temperature-controlled crystal oscillators (TCXOs), 296
- Terrain-aided inertial navigator (TERRAIN), 191–92
 - general positioning system case study, 202–5
 - horizontal/vertical guidance, 193
 - LADAR-based integration
 - methodology, 192–95
 - mode, 194
 - position estimator, 195
 - precision approach guidance system
 - case study, 205–10
 - precision-approach hardware diagram, 206
 - proof-of-concept precision approach system, 208
 - real-time precision approaches, 205
 - system performance case studies, 202–10
- The time to first fix (TTFF), 144
- Tight GNSS/INS integration, 162–63
 - benefits, 162
 - defined, 162
 - disadvantage, 162
 - illustrated, 163
 - tracking loop aiding, 163
 - See also* GNSS/INS integration
- Time
 - estimating, 66–67
 - GPS, 66
 - update, 166
 - See also* Position, velocity, and time (PVT)

- Time difference-of-arrival (TDOA)
 - systems, 214
 - advantage, 226
 - arrival times, 225
 - location from, 225
- Time domain derivation, 476–78
- Time DOP (TDOP), 63
- Time multiplex method of combination (TMBOC), 454, 457–59
 - defined, 457
 - spreading symbols, 458
- Time-of-arrival (TOA) systems, 214
 - location from, 224
 - measurements from UWB signals, 226
 - TDOA advantage over, 226
- Time of transmission (TOT), 274
- Time-to-alert (TTA), 253, 263–64
- Timing (phase), 213, 223–26
- Tracking
 - BOC, 44–47
 - carrier phase (CPH), 365
 - closed-loop (CL), 392
 - frequency-locked loop (FLL), 34
 - GPS software receiver, 136–41
 - multiple hypothesis (MHT), 233
 - PSK signal, 34–39
 - weak signals, 296–97
- Tracking loops, 138–39
 - aiding, 163
 - domain integration, 287
 - gains, 129
- Transmit power levels, 11–12
- Transport rate, 155
- Trapezoidal sine function, 50
- Triangular cosine function, 40
- Triangulation
 - angulation, 227–28
 - combined lateration and angulation, 228–29
 - defined, 226
 - deterministic estimation with, 226–29
 - lateration, 226
- Triple Carrier Ambiguity Resolution (TCAR), 376
- Triple-differencing mode, 354
- Troposphere term, 68
- Tropospheric spatial-decorrelation anomalies, 114
- Trs () function, 49
- TV-based location, 238–39
- U
 - Ultra stable oscillators (USOs), 384
 - Unmanned aerial vehicles (UAVs), 149
 - Unscented Kalman filter (UKF), 165
 - User range accuracy (URA), 105
 - User range errors (UREs), 103–4
- V
 - Van-Cittert-Zernike theorem, 414
 - Vector DLL (VDLL), 164
 - Vehicle sideslip estimation, 180–86
 - categories, 181
 - defined, 180–81
 - with filter-blending, 183
 - hypothetical vehicle trajectories, 184
 - lateral force estimates, 183
 - motivation, 181–83
 - observability, 184–85
 - sideslip angle, 181, 182
 - See also* GNSS/INS integration
 - Velocity update equation, 155–56
 - Vertical DOP (VDOP), 63, 64
 - VHF omnidirectional range (VOR), 218, 248
 - bearing estimation receiver, 219
 - GNSS combination with, 218
 - instrumentation, 219
- W
 - Wave sensing, 423–24
 - Weak signal navigation, 7, 291–324
 - acquisition example, 319–22
 - aiding and assistance possibilities, 291–92
 - algorithms, 303–6
 - assistance, 300–301
 - computational aspects, 292
 - cross-correlation, 293, 297
 - DVD examples, 319–23
 - fault detection and exclusion methods, 292
 - FDE example, 322–23
 - hardware issues, 293

- Weak signal navigation (continued)
 - integration, 293
 - integrity monitoring, 306–16
 - multipath, 293
 - quality control, 316–18
 - signal processing issues, 291, 292–300
 - supportive systems, 301–3
- Weak signals
 - acquisition of, 294–95
 - clock stability and, 295–96
 - RAIM and, 306
 - tracking, 296–97
 - varying environment and, 306
- Weighted least-squares notation, 308–10
 - estimate, 309
 - positioning result, 323, 324
 - solution, 310
- Weiner-Khinchine Theorem, 444
- WGS84 standard, 57
 - position coordinates, 58
 - reference ellipsoid, 156
- Wide-Area Augmentation System (WAAS), 94
 - HAT for, 209
 - inertial Kalman filter, 194
 - ionosphere corrections, 129
 - Performance Analysis Network (PAN), 94
- Wide-area DGNSS, 91, 93–94
 - defined, 93
 - local-area DGNSS versus, 93–94
 - See also* Differential GNSS (DGNSS)
- WiFi, 217
- WLAN systems, 302
- Y**
 - YARD HOUND system, 215, 216
- Z**
 - ZigBee, 217

The GNSS Technology and Applications Series

Elliott Kaplan and Christopher Hegarty, Series Editors

A-GPS: Assisted GPS, GNSS, and SBAS, Frank van Diggelen

Applied Satellite Navigation Using GPS, GALILEO, and Augmentation Systems,
Ramjee Prasad and Marina Ruggieri

Digital Terrain Modeling: Acquisition, Manipulation, and Applications, Naser
El-Sheimy, Caterina Valeo, and Ayman Habib

Geographical Information Systems Demystified, Stephen R. Galati

GNSS Applications and Methods, Scott Gleason and Demoz Gebre-Egziabher

GNSS Markets and Applications, Len Jacobson

GNSS Receivers for Weak Signals, Nesreen I. Ziedan

Introduction to GPS: The Global Positioning System, Second Edition,
Ahmed El-Rabbany

Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems,
Paul D. Groves

Spread Spectrum Systems for GNSS and Wireless Communications, Jack K. Holmes

Understanding GPS: Principles and Applications, Second Edition, Elliott Kaplan and
Christopher Hegarty, editors

Ubiquitous Positioning, Robin Mannings

Wireless Positioning Technologies and Applications, Alan Bensky

For further information on these and other Artech House titles,
including previously considered out-of-print books now available through our
In-Print-Forever® (IPF®) program, contact:

Artech House Publishers

685 Canton Street

Norwood, MA 02062

Phone: 781-769-9750

Fax: 781-769-6334

e-mail: artech@artechhouse.com

Artech House Books

46 Gillingham Street

London SW1V 1AH UK

Phone: +44 (0)20 7596 8750

Fax: +44 (0)20 7630 0166

e-mail: artech-uk@artechhouse.com

Find us on the World Wide Web at: www.artechhouse.com